



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

Centro de Ciencias Genómicas

Modificando la red de regulación transcripcional para la reducción del proteoma no asociado al crecimiento.

TESIS

que para optar el grado de:

DOCTOR EN CIENCIAS

presenta:

Gustavo Moisés Lastiri Pancardo

Tutor Principal:

Dr. José Utrilla Carreri

Centro de Ciencias Genómicas

Cuernavaca, Morelos, noviembre de 2020



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

La presente tesis se realizó en el programa de biología sintética y de sistemas del Centro de Ciencias Genómicas de la UNAM, con el apoyo de la beca de manutención CONACYT #434655 a nivel doctorado, los proyectos UNAM–DGAPA-PAPIIT IA200716 y IA201518 además de Newton advanced Fellowship Project NA 160328.

Agradecimientos

A mis padres, amigos, compañeros y colegas

Resumen

El control de la asignación de recursos celulares en sistemas biológicos para aplicaciones de biología sintética es uno de los principales desafíos de esta disciplina. Los organismos silvestres invierten una gran cantidad de recursos celulares para garantizar la supervivencia en entornos cambiantes, reduciendo la eficiencia de las funciones sintéticas introducidas. En el presente trabajo se presenta una metodología novedosa para controlar la asignación de recursos en *Escherichia coli* modificando de manera racional la red de regulación transcripcional (TRN). El método (ReProMin) identifica el conjunto mínimo de intervenciones genéticas que maximiza el ahorro en recursos celulares que normalmente se utilizarían para expresar genes no esenciales. Primero se categorizaron los factores de transcripción (TF) según la esencialidad de los genes que regulan, después se integraron datos proteómicos disponibles para clasificarlos en función de su balance proteómico (PB), definido como la masa proteómica neta que liberan. Realizando un análisis combinatorio, se buscó la combinación de TFs que maximiza la liberación de masa proteómica y se validaron las predicciones del método experimentalmente. La cepa diseñada resultante contiene sólo tres mutaciones y su perfil de expresión transcripcional mostró que las intervenciones regulatorias hechas son altamente específicas; dichas mutaciones liberan teóricamente el 0.5% de su proteoma. La cepa posee un presupuesto proteómico más alto y exhibió un mayor rendimiento de producción de un metabolito generado a partir de una vía metabólica recombinante. La metodología presentada en este trabajo muestra que el combinar datos proteómicos y de regulación de células completas es una forma efectiva de optimizar fenotipos de una manera predecible utilizando métodos moleculares convencionales.

Importancia del trabajo

La capacidad para diseñar sistemas biológicos depende de comprender cómo las células perciben y responden a su entorno a nivel sistema, pocos estudios han abordado este tema y ninguno de ellos de manera racional. Se prevé que haciendo uso de los mecanismos reguladores de la transcripción facilitará el diseño de fenotipos celulares.

En este trabajo se desarrolló una nueva estrategia de ingeniería celular para *E. coli* utilizando la TRN como control para la asignación de recursos proteómicos. Al combinar información proteómica de alto rendimiento, interacciones de redes reguladoras y observaciones de esencialidad genética, se desarrolló un método capaz de encontrar el conjunto mínimo de intervenciones genéticas necesarias para desviar los recursos invertidos en funciones de cobertura hacia funciones sintéticas.

Índice General

| | |
|--|----|
| Agradecimientos..... | 2 |
| Resumen..... | 3 |
| Importancia del trabajo..... | 3 |
| Índice General..... | 4 |
| Índice de Figuras..... | 6 |
| Índice de Tablas..... | 7 |
| Abreviaturas..... | 8 |
| Introducción..... | 9 |
| 1.1 Biología sintética..... | 9 |
| 1.2 Biología de sistemas..... | 9 |
| 1.3 Regulación transcripcional..... | 10 |
| 1.4 Red de Regulación Transcripcional..... | 11 |
| 1.5 Funciones de cobertura..... | 11 |
| 1.6 Modelos metabólicos a escala genómica (GEM)..... | 12 |
| 1.7 Estrategias para el diseño de fenotipos microbianos..... | 12 |
| Objetivos..... | 13 |
| 2.1 Planteamiento del problema..... | 13 |
| 2.2 Antecedentes..... | 13 |
| 2.3 Hipótesis..... | 15 |
| 2.4 Objetivo general..... | 15 |
| 2.5 Objetivos específicos..... | 15 |
| Resultados..... | 16 |
| 3.1 Simulaciones con un modelo ME..... | 16 |
| 3.2 Identificación de TFs dispensables para la liberación de proteoma en una condición definida..... | 17 |
| 3.3 Integración de datos proteómicos y de regulación..... | 18 |
| 3.4 Búsqueda computacional de combinaciones de TFs para maximizar liberación de recursos..... | 19 |
| 3.5 Los cálculos de ReProMin revelaron el potencial de liberación del proteoma..... | 20 |

| | |
|---|----|
| 3.6 El análisis de RNA-seq confirmó la alta especificidad de las mutaciones introducidas..... | 25 |
| 3.7 Cuantificación de la masa del proteoma liberado..... | 28 |
| 3.8 Evaluación fenotípica de las cepas diseñadas..... | 29 |
| 3.9 Expresión de una vía metabólica sintética: producción de violaceína. | 35 |
| Discusión..... | 39 |
| Perspectivas..... | 42 |
| Métodos | 44 |
| Apéndice | 50 |
| Bibliografía | 86 |

Índice de Figuras

| | |
|---|----|
| Figura 1. Simulaciones del modelo ME y respuesta de los sectores del proteoma a la reducción de la fracción de proteína no utilizada (UPF)..... | 17 |
| Figura 2. Esencialidad genética y análisis de la red de regulación transcripcional en la condición predefinida..... | 19 |
| Figura 3. Propiedades emergentes derivadas de la integración de datos proteómicos..... | 20 |
| Figura 4. Cálculos de liberación del proteoma usando ReProMin..... | 22 |
| Figura 5. Espacios de liberación del proteoma calculados por ReProMin correspondientes al caso de UT en distintas condiciones de crecimiento..... | 23 |
| Figura 6. Representación gráfica de las predicciones de ReProMin de las mutantes generadas..... | 25 |
| Figura 7. Análisis transcriptómico de la cepa diseñada para el caso UT..... | 28 |
| Figura 8. Estimación de la masa de la fracción del proteoma asociada a los TFs eliminados..... | 30 |
| Figura 9. Evaluación fenotípica de las cepas basada en las predicciones de ReProMin para glucosa (casos de UT y ST) y el control..... | 31 |
| Figura 10. Evaluación de la carga metabólica de las cepas basadas en las predicciones de ReProMin para medio mínimo con glucosa (casos UT y ST) y el control..... | 32 |
| Figura 11. Caracterización del circuito sintético..... | 33 |
| Figura 12. Líneas de isocosto durante el crecimiento equilibrado (~5 h)..... | 35 |
| Figura 13. Líneas de isocosto durante fase estacionaria (~22 h)..... | 36 |
| Figura 14. Evaluación de la vía de la violaceina..... | 37 |
| Figura 15. Caracterización de la producción de violaceina..... | 38 |

Índice de Tablas

| | |
|--|----|
| Tabla 1. Lista de genes esenciales definidos para el crecimiento en medio M9 glucosa..... | 51 |
| Tabla 2. Información de los TFs candidatos definidos por ReProMin..... | 59 |
| Tabla 3. Resumen del análisis combinatorio ReProMin de los casos de ST y UT.... | 60 |
| Tabla 4. Datos RNA-seq y proteómicos estimados para los objetivos ReProMin de PFC y los off-targets..... | 61 |
| Tabla 5. Cepas, Primers y Plásmidos utilizados en el estudio..... | 64 |

Abreviaturas

Factor de Transcripción (**TF**)

Red de Regulación Transcripcional (**TRN**)

Regulón Densamente Sobrelapado (**DOR**)

Módulo de entrada única (**SIM**)

Minimización del proteoma basado en regulación (**ReProMin**)

Capítulo 1

Introducción

1.1 Biología sintética

La biología sintética (SynBio) es un área de investigación multidisciplinaria que combina enfoques de ingeniería y computación donde se busca crear nuevas partes, dispositivos y sistemas biológicos, o rediseñar sistemas que ya se encuentran en la naturaleza con el objetivo de dotarlos de características superiores o funciones totalmente nuevas para así generar compuestos, sistemas o materiales que puedan mejorar la condición humana¹

De manera general las estrategias empleadas en SynBio se pueden dividir en dos enfoques diferentes: “top-down” y “bottom-up”.

Top-down. Este enfoque implica el uso de técnicas de ingeniería metabólica y genética para impartir nuevas funciones a sistemas biológicos ya existentes.

Bottom-up. Implica la creación de nuevos sistemas biológicos *in vitro* mediante la unión de componentes biomoleculares "no vivos", con el objetivo de construir una célula artificial.

Actualmente la biología sintética está creciendo rápidamente debido a las nuevas y potentes técnicas de ingeniería genética y a la disminución de los costos de síntesis y secuenciación de ADN, por lo que se espera que pronto aplicaciones basadas en biocómputo, biosensores, proteínas de diseño, biomateriales, plataformas para entrega y administración de fármacos, sistemas libres de células, entre otras, sean más comunes y baratas.

1.2 Biología de sistemas

La biología de sistemas es una disciplina científica en la que se utilizan modelos matemáticos y computacionales para estudiar los sistemas biológicos. Estos sistemas típicamente implican complejas redes metabólicas o de señalización. La biología de sistemas utiliza un enfoque holístico que estudia el sistema biológico completo, a diferencia de la biología molecular que se centra en subsistemas que a menudo se estudian mediante experimentos independientes. Una característica particular de esta disciplina es que implica un análisis cuantitativo, a diferencia de la naturaleza mayormente cualitativa de otros campos, que se centra en la prueba de hipótesis donde sólo se describe si el sistema es verdadero o falso. A pesar de los diferentes enfoques, la biología de sistemas depende en gran medida de la

amplia información biológica que se ha adquirido a través de biología molecular, y los estudios de biología de sistemas a menudo resultan en la generación de hipótesis que requieren una confirmación mediante un enfoque reduccionista.

Al igual que en SynBio, en la biología de sistemas también existen dos enfoques diferentes de estudio.

Top-down. Donde distintos tipos de datos “omicos” de alto rendimiento a diferentes niveles (genómico, transcriptómico, proteómico y metabolómico) se analizan de forma integradora y relacional.

Bottom-up. Se busca desarrollar modelos detallados para procesos específicos, por ejemplo, las reacciones enzimáticas individuales de un organismo se ensamblan en un modelo que describe el sistema que se estudia.

Los dos enfoques son complementarios; mientras que las estrategias “top-down” son útiles para trazar el mapa de las funciones celulares a escala del genoma, el enfoque “bottom-up” permite una escala temporal detallada del impacto de los componentes individuales en las propiedades generales del sistema. Al complementarse, ambos enfoques permiten la identificación de lo que se conoce como propiedades emergentes del sistema, es decir, cómo las interacciones entre los diferentes componentes celulares dan lugar a funciones biológicas que no pueden ser identificadas mediante un enfoque reduccionista.

1.3 Regulación transcripcional

El ambiente externo e interno de los sistemas biológicos está en constante cambio por lo que muchas de sus funciones sólo son necesarias bajo ciertas condiciones. Por ejemplo, producir las proteínas necesarias para metabolizar una fuente determinada de carbono sólo será necesario cuando dicha fuente se encuentre en el medio. Producir constantemente todas las proteínas codificadas en el genoma, incluso en momentos en los que no son necesarias, representaría un gasto importante que podría comprometer la viabilidad de las células, así que para evitar este compromiso las células son capaces de controlar la expresión de grupos de genes según el ambiente en el que se encuentren. Una de las principales formas de regulación genética es la regulación transcripcional donde se busca promover o inhibir la transcripción de genes.

La forma más común de regulación transcripcional en bacterias está a cargo de proteínas llamadas Factores de Transcripción (TFs). Los TFs son capaces de unirse a secuencias específicas en el DNA llamados promotores los cuales generalmente se encuentran en regiones cercanas al sitio de inicio de la transcripción, al unirse al

DNA los TFs son capaces de promover o inhibir la transcripción de los genes subsecuentes. De acuerdo a su actividad reguladora, los TFs pueden ser:

Activadores. TFs que promueven la transcripción de sus genes regulados.

Represores. TFs que inhiben la transcripción de sus genes regulados.

Duales. TFs que son capaces de promover o impedir la transcripción de sus genes regulados según la presencia de señales ambientales.

1.4 Red de Regulación Transcripcional

Dado que un TF puede regular varios genes y cada gen puede ser regulado por más de un TF, es posible representar al conjunto de interacciones de regulación en forma de red, donde los nodos representan TFs o genes y las aristas representan interacciones de regulación. A dicha red se le denomina Red de Regulación Transcripcional (TRN) por sus siglas en inglés.

Las redes de regulación han sido sujetas a múltiples estudios que han permitido dilucidar las propiedades biológicas y matemáticas que rigen su estructura y evolución. Los organismos responden al medio ambiente mediante vías de señalización celular codificadas en sus redes regulatorias² y la complejidad del estilo de vida de un organismo generalmente se traduce en la complejidad de la señalización³. Las redes de regulación transcripcional son robustas⁴ y evolucionan⁵ para hacer frente a las perturbaciones ambientales y de estilo de vida particular de cada organismo, sin embargo, esta solidez implica compromisos intrínsecos en la dinámica de asignación de recursos celulares.

1.5 Funciones de cobertura

Los organismos silvestres necesitan asignar recursos celulares a las funciones necesarias para su crecimiento y supervivencia de una manera dinámica. Se ha observado que, para maximizar la supervivencia en su ambiente natural, los sistemas biológicos están evolutivamente "pre-programados" para anticiparse a fluctuaciones en el medio ambiente (por ejemplo, *Escherichia coli* típicamente se enfrenta a un choque ácido cuando pasa a través del estómago de un mamífero)^{6,7}.

Las funciones que preparan a un organismo para un cambio ambiental o un estrés próximo, se denominan funciones de cobertura y su mecanismo de expresión ha sido ampliamente documentado^{8,9}. La expresión de dichas funciones está codificada en la red reguladora y tiene un costo proteómico, drenando recursos de otras

funciones prioritarias¹⁰. Las funciones de cobertura son innecesarias en un ambiente de laboratorio típicamente estable.

1.6 Modelos metabólicos a escala genómica (GEM)

Los modelos metabólicos a escala genómica (GEM) son una herramienta computacional muy útil que permite describir a la vez todo el metabolismo de un organismo utilizando información genómica. Un GEM describe exhaustivamente todas las reacciones y vías metabólicas de un organismo utilizando coeficientes estequiométricos de los metabolitos participantes¹¹.

Estos modelos han tenido un enorme impacto en la biología de sistemas y la ingeniería metabólica ya que permiten calcular el proteoma teórico mínimo y el costo celular de expresar funciones de cobertura en una condición específica de crecimiento¹².

1.7 Estrategias para el diseño de fenotipos microbianos

La eliminación de las funciones accesorias no esenciales es una de las estrategias utilizadas para diseñar fenotipos microbianos optimizados. Este enfoque se basa en la suposición de que los recursos celulares para la expresión genética son limitados y, por lo tanto, al eliminar genes innecesarios en un determinado entorno, la célula es capaz de reasignar los recursos liberados a otras funciones (por ejemplo, expresión de genes recombinantes). Muchos intentos de optimización se realizan principalmente reduciendo el tamaño del genoma y el número de genes, frecuentemente de manera masiva y aleatoria^{13,14}; sin embargo, estos enfoques no consideran la forma precisa en que se realiza la reasignación de recursos después de las intervenciones genéticas.

Los enfoques econométricos para estudiar la asignación de recursos celulares pueden facilitar el desarrollo de fenotipos destinados a mostrar una función de sintética. Se han realizado diversos estudios en *E. coli* sobre la asignación de recursos celulares en los cuales se ha dividido el proteoma en sectores funcionales (análisis “coarsed-grain”), también se ha logrado establecer los principios “econométricos” que rigen dichos sectores y como estos se ven afectados por la regulación genética^{15,16}, además mediante el uso de modelos a escala genómica es posible calcular los recursos celulares necesarios para sostener el crecimiento en una condición en particular y expresar una función sintética simultáneamente¹².

Los organismos sintéticos obtenidos utilizando modelos econométricos tienen menos probabilidades de desarrollar comportamientos emergentes no deseados¹⁷.

Capítulo 2

Objetivos

2.1 Planteamiento del problema

Mientras que la capacidad de auto-organización y auto-regulación son dos de las principales ventajas que presentan los sistemas biológicos en comparación con los sistemas artificiales, esto a la vez implica una alta complejidad en la estructura de organización y operación de estos sistemas, lo que complica su estudio.

Los organismos que hoy en día se emplean en aplicaciones de SynBio conservan los programas regulatorios nativos los cuales ayudan a maximizar su supervivencia en ambientes naturales típicamente variables, lo anterior trae como consecuencia una alta variabilidad en la respuesta de los circuitos sintéticos introducidos en condiciones estables de laboratorio y por lo tanto una baja capacidad de estandarización de los componentes empleados. Además, el desempeño y la eficiencia no es suficiente para satisfacer las necesidades que se requieren a escala industrial, es por esto que hasta el momento los logros de esta disciplina se han restringido a sistemas muy sencillos y de pocos componentes.

Desde un punto de vista ingenieril, el campo necesita la generación de organismos en los cuales se puedan integrar circuitos sintéticos con partes estandarizables, reproducibles y escalables. Actualmente se están llevando a cabo varios esfuerzos para superar estas limitaciones, tales como la reducción del tamaño del genoma, el diseño de un organismo mínimo por medio del ensamblado de genomas sintéticos¹³ o la expresión de funciones sintéticas empleando sistemas ortogonales (que no interactúan con los componentes del organismo huésped)¹⁸, lo que ha resultado en avances considerables, sin embargo, pocos esfuerzos se están haciendo para optimizar y controlar la asignación de recursos celulares en el huésped para invertir en funciones sintéticas, por otro lado, también es necesario desarrollar metodologías sencillas que puedan ser fácilmente extrapolables a diversos organismos.

2.2 Antecedentes

Hoy en día para *E. coli* existen una gran cantidad de bases de datos “omicos” con un sin fin de información sobre prácticamente cualquier proceso celular, también se cuenta con modelos ME que simulan la mayoría de los procesos metabólicos y de expresión genética en una célula y que permiten cuantificar las demandas celulares

de expresión proteómica para el crecimiento en un ambiente nutricional determinado¹⁹.

Recientemente se han realizado estudios donde comparando predicciones de un modelo ME contra datos experimentales se clasificó el proteoma de *E. coli* en sectores funcionales basados en función celular. Derivado de este análisis, se logró identificar fracciones del proteoma expresado, pero no usado y fracciones del proteoma subutilizado. Una de las principales conclusiones de estos estudios es que la variabilidad de la velocidad de crecimiento en ambientes nutricionales similares está dada por la expresión de fracciones del proteoma no usados que generan fenotipos sub-óptimos^{10,20}.

Estos estudios resaltan el hecho de que el tamaño del proteoma es limitado y que la introducción de funciones exógenas sintéticas ejerce una carga sobre las funciones celulares endógenas de por sí ya limitadas, además señalan que el sector proteómico de las funciones de cobertura no se utiliza bajo condiciones de laboratorio estables y que por el contrario su expresión reduce el crecimiento considerablemente. También se observó que la reducción en la expresión del 2 - 5% de los genes asociados a funciones de cobertura puede redirigir alrededor de un tercio de la energía de estas funciones hacia crecimiento; llevando a un aumento del 22% en la velocidad de crecimiento en un entorno constante¹⁰.

Experimentos de evolución adaptativa (ALE) han demostrado un gran éxito, especialmente para identificar funciones no relacionadas con el crecimiento²¹, sin embargo, esta metodología es empleada para seleccionar cepas optimizadas en crecimiento rápido que no necesariamente resultan en los mejores fenotipos de producción. Además, los cambios genéticos generados en ALE normalmente son complicados de explicar, por lo tanto, sus efectos no son predecibles²².

Tomando en cuenta las observaciones anteriores se ha planteado la idea de que al reducir la expresión de la fracción del proteoma que no está asociada al crecimiento se podría aumentar la disponibilidad de recursos celulares para así mejorar la reproducibilidad y el rendimiento de los circuitos sintéticos introducidos. En el presente trabajo se hará uso del conocimiento sobre los principios econométricos del proteoma, junto al de la red de la TRN para llevar a cabo la reducción del proteoma no asociado al crecimiento con el objetivo de incrementar el rendimiento de un circuito sintético en un huésped bacteriano.

2.3 Hipótesis

Al modificar la TRN de *E. coli* se logrará reducir la expresión de proteínas no asociadas al crecimiento aumentando la disponibilidad de recursos celulares para la expresión de funciones sintéticas.

2.4 Objetivo general

Desarrollar una metodología que permita modificar la TRN de *E. coli* para llevar a cabo la reducción del proteoma de cobertura.

2.5 Objetivos específicos

- Identificar los genes asociados a la expresión de funciones de cobertura en un ambiente específico.
- Analizar la TRN en busca de los TFs involucrados en la expresión de los genes de cobertura.
- Integrar datos “omicos” que permitan cuantificar la cantidad de recursos celulares controlada por cada TF.
- Realizar un análisis que permita encontrar la combinación de TFs que maximiza la cantidad de recursos liberados.
- Construir cepas knock-out basadas en el análisis previo.
- Evaluar los efectos transcripcionales de las mutaciones introducidas.
- Evaluar el desempeño fenotípico de las cepas construidas.
- Evaluar la aplicación de las cepas mutantes como posible organismo chasis en biología sintética.

Capítulo 3

Resultados

3.1 Simulaciones con un modelo ME.

Se utilizó el modelo ME iJL1678b-ME para simular diferentes escenarios de expresión del proteoma de cobertura. De manera similar al coeficiente de energía de mantenimiento, el proteoma de cobertura y otras funciones no relacionadas con el crecimiento (por lo tanto, no modeladas) se tienen en cuenta en los modelos ME como parte del coeficiente de la fracción de proteína no utilizada (UPF) (ver métodos)²³. Lo anterior se realizó variando en el modelo el parámetro asociado a la UPF, para simular el efecto de la reducción de la UPF y diferentes niveles de expresión de una proteína modelo recombinante no utilizada (GFP).

Las simulaciones realizadas permitieron calcular los costos y beneficios de diferentes intervenciones genéticas, modulando la expresión del proteoma de cobertura, expresada en términos de crecimiento y el tamaño de los sectores del proteoma esencial y recombinante (Figura 1). Tomando como base las observaciones derivadas de estas simulaciones, se busca diseñar cepas conteniendo las mínimas intervenciones genéticas que reduzcan la mayor cantidad de recursos proteómicos que no sean requeridos para crecer en una condición específica.

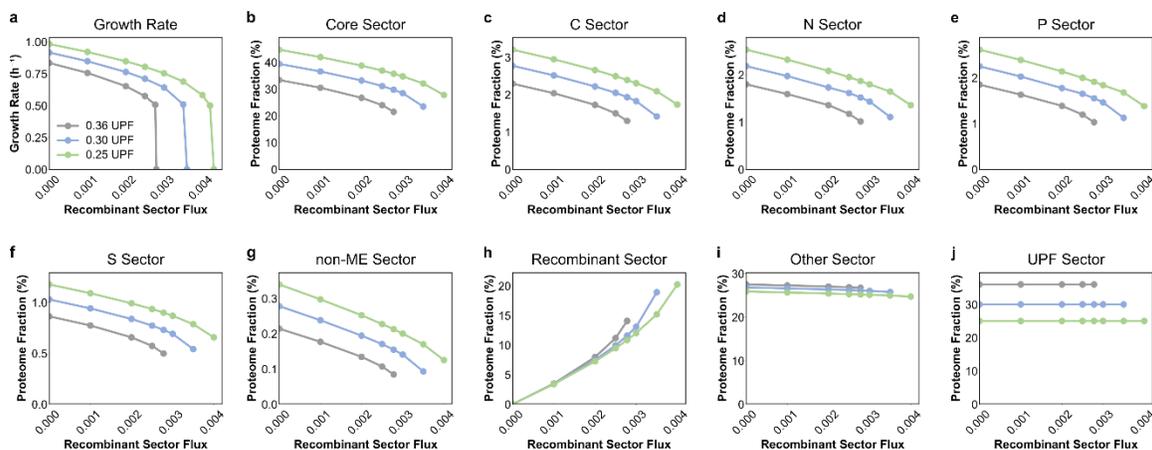


Figura 1 | Simulaciones del modelo ME y respuesta de los sectores del proteoma a la reducción de la fracción de proteína no utilizada (UPF). Cada panel muestra a, la tasa de crecimiento y la fracción correspondiente de cada sector del proteoma b, sector core y el sector dependiente de elementos alternativos: c, el sector de carbono d, el sector de nitrógeno e, el sector de fósforo f, el sector de azufre g, el sector no ME h, el sector

recombinante, comprendido por la expresión máxima alcanzada de la GFP, i, el sector (no clasificado) y j, el sector UPF.

3.2 Identificación de TFs dispensables para la liberación de proteoma en una condición definida.

El principio del método consiste en utilizar TFs como interruptores que controlan la asignación del proteoma de cobertura en un entorno específico predefinido. Debido a que es una condición de crecimiento muy común y bien documentada, se decidió establecer como condición el crecimiento en lote en medio mínimo (M9) suplementado con glucosa como la única fuente de carbono.

Se comenzó por generar una lista de genes esenciales para el crecimiento en este entorno específico, para esto se combinaron cinco distintos sets de datos con observaciones de esencialidad de genes a escala genómica generados *in vivo* e *in silico*. (Figura 2a, Tabla 1, ver métodos).

Una vez que se definió la esencialidad genética específica del caso, se integró la información sobre las interacciones TF-gen depositadas en RegulonDB²⁴. Se examinó la subred de interacciones de cada TF en busca de aquellos que son prescindibles, los cuales se definen como aquellos que no activan la expresión de ningún gen esencial. Según el análisis, se pueden eliminar 156 de los 200 TF contenidos en la red reguladora (Figura 2b).

Dado que el objetivo es reducir el proteoma de cobertura, de los 156 TFs prescindibles, se seleccionaron como candidatos 34 TFs con al menos un gen único regulado positivamente (lo que significa que el gen no es activado por ningún otro TF) (Tabla 2, ver métodos); lo anterior da la certeza de silenciar al menos un gen al eliminar el TF.

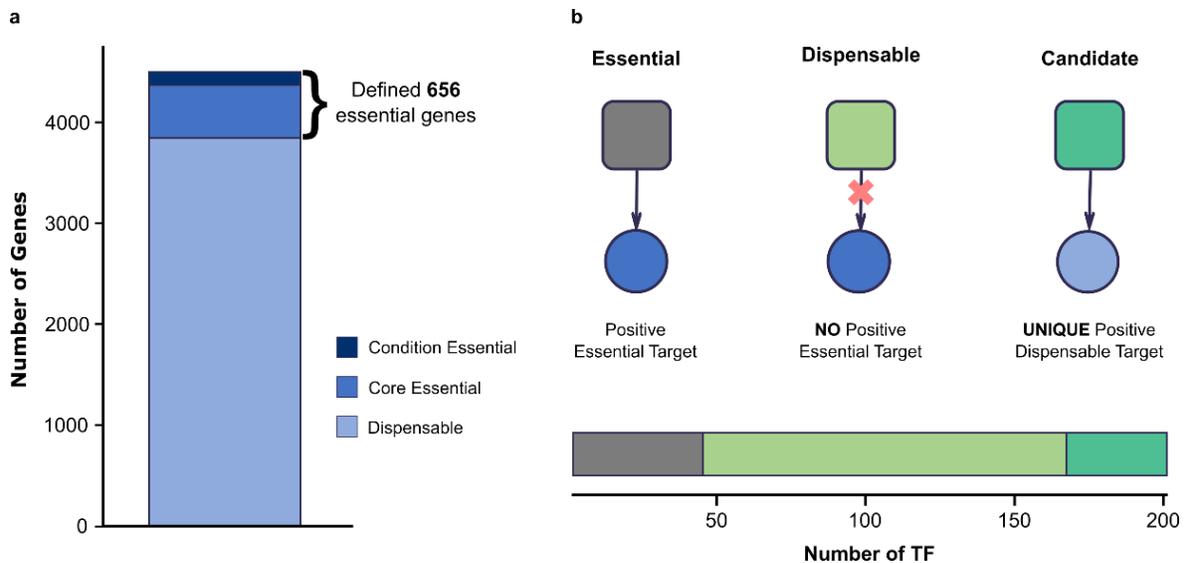


Figura 2 | Esencialidad genética y análisis de la red de regulación transcripcional en la condición predefinida. **a**, Perfil de esencialidad de los genes de *E. coli* (~4.600) en la condición de crecimiento seleccionada. Los genes esenciales considerados para este análisis se dividen en esenciales core (523 genes: siempre necesarios) y condicionales (133 genes: necesarios sólo en M9-glucosa). **b**, Representación gráfica de la subred de interacciones TF-gen considerada para la clasificación de los TFs; los cuadrados grises representan los TFs esenciales, los cuadrados verde claro los TFs prescindibles, los cuadrados verde oscuro los TFs candidatos, los círculos azul oscuro los genes esenciales, los círculos azul claro los genes prescindibles y las flechas las interacciones positivas. La barra inferior muestra el número de TFs en cada clasificación del total considerado (203 TFs).

3.3 Integración de datos proteómicos y de regulación.

Para determinar el tamaño de la fracción del proteoma asociado a cada TF, se hizo uso de un conjunto de datos proteómicos cuantitativos²⁵ el cual proporciona el número de copias de proteínas por célula en 22 condiciones de crecimiento diferentes con una cobertura del 95% del proteoma (en masa).

En consecuencia, es posible definir dos propiedades emergentes derivadas de la integración de los datos proteómicos cuantitativos: la carga proteómica de un gen (PL) en fg de proteína por célula (Figura 3a) y el balance proteómico (PB) de un TF resultante de la suma del PL de los genes que resultarían silenciados o activados por la eliminación del TF (Figura 3b). El PB es conceptualmente importante para clasificar los TF según el tamaño del proteoma que controlan, ya que considera la masa neta (en fg de proteína/célula) liberada al eliminar un TF.

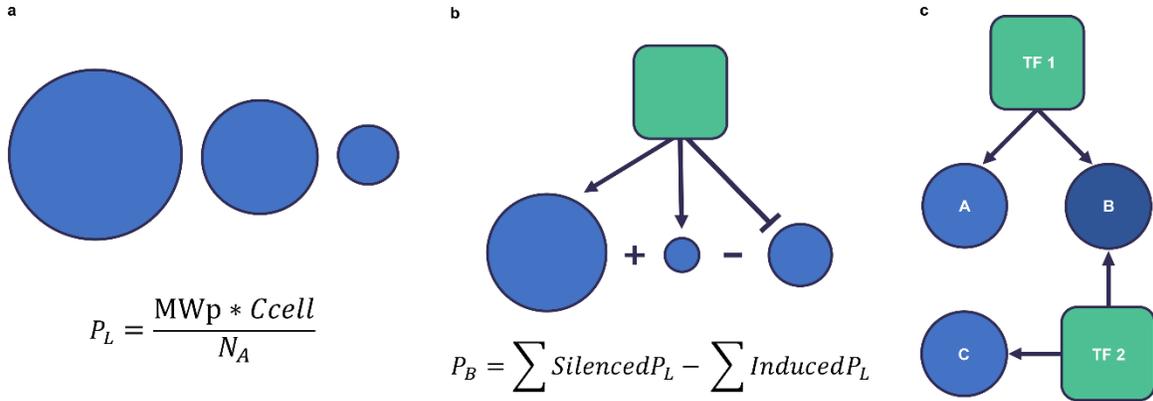


Figura 3 | Propiedades emergentes derivadas de la integración de datos proteómicos.

a, La carga proteómica de un gen (PL) se define como el peso molecular de la proteína (MWp) multiplicado por el número de copias por célula (Ccell) dividido por el número de Avogadro (N_A) (6.022×10^8 equivalente a fg). Cuanto más expresado está el gen, más carga proteómica genera. **b**, El balance proteómico de un TF (PB) definido como la suma de la PL de los genes silenciados menos la suma de la PL de los genes inducidos. **c**, Esquema de un caso simple de regulación compartida en el que al eliminar ambos TF se silencian todos los blancos, pero no es así cuando los TF se eliminan individualmente.

3.4 Búsqueda computacional de combinaciones de TFs para maximizar liberación de recursos.

Muchos TF poseen genes blanco compartidos (Figura 3c), de hecho, varios de ellos forman parte de una versión simplificada de un motivo de redes llamado “Dense Overlapping Regulon” (DOR)²⁶, lo que significa que teóricamente se necesita una combinación particular de TFs para tener la certeza de que estos genes están silenciados por completo, lo que genera un espacio de potencial liberación de proteoma, el cual está compuesto por todas las posibles combinaciones de TFs.

Para ayudar en el diseño de las cepas mutantes se desarrolló un método de cálculo; denominado ReProMin (Regulation based Proteome Minimization – minimización del proteoma basada en regulación); el método integra la red de interacción TF-gen y los datos proteómicos cuantitativos para generar y resolver el espacio de liberación de proteoma para una lista particular de TFs, devolviendo finalmente la combinación de mutaciones de tamaño N que silencia la mayor carga proteómica en una condición de crecimiento específica (ver métodos).

Para una lista de TFs, los límites del espacio están definidos por la combinación que libera el PL más alto y más bajo para todos los tamaños de combinación de TF posibles; lo que significa que cuanto más grande sea la lista de TF, más grande y complejo será el espacio de liberación. El tiempo necesario para resolver el espacio

depende del número total de combinaciones, que aumenta exponencialmente a medida que la lista de TF se hace más grande (ver ecuación 2 en los métodos)

3.5 Los cálculos de ReProMin revelaron el potencial de liberación del proteoma.

Se utilizó la condición de medio mínimo de glucosa como punto de partida para el análisis. Se consideraron dos casos distintos para realizar los cálculos. El caso de blancos compartidos (ST): considerando todos los TFs no esenciales con PB positivo (132 TF), en este caso se tienen en cuenta algunos TFs sin genes únicos regulados y el caso de blancos únicos (UT): considerando los TF candidatos previamente definidos con un PB positivo (20 TF).

Para el caso ST la herramienta de cálculo fue capaz de resolver combinaciones de hasta seis TFs ($>6 \times 10^9$ combinaciones, tomando aproximadamente 95 h de cómputo, Tabla 3), el siguiente caso fue imposible resolverlo debido al número extremadamente grande de combinaciones (siete TFs $>11 \times 10^{10}$ combinaciones). Para maximizar el espacio resuelto, se continuó resolviendo otras áreas del paisaje donde el número de combinaciones es lo suficientemente pequeño como para ser resuelto (cuando $n \rightarrow 0$ ó $n \rightarrow r$)

Los cálculos revelaron que la eliminación de todos los 132 TF no esenciales no es el mejor caso de liberación del proteoma. El máximo se alcanza con la eliminación de 130 TFs donde potencialmente se liberaría hasta el 1.06% del proteoma total de *E. coli*, por otro lado, hasta el 0.94% puede ser liberado eliminando la mejor combinación de seis TFs, mientras que hasta el 0,53% del proteoma total puede ser liberado eliminando sólo la mejor combinación de tres TFs (Figura 4a).

Para el caso UT, la eliminación de todos los 20 TFs candidatos liberaría hasta el 0.72% del proteoma, sin embargo, los cálculos muestran que no hay una mejora significativa en la liberación de recursos después de la eliminación de ocho TFs. De hecho, el 60% de la potencial liberación puede lograrse eliminando sólo tres TFs (Figura 4b).

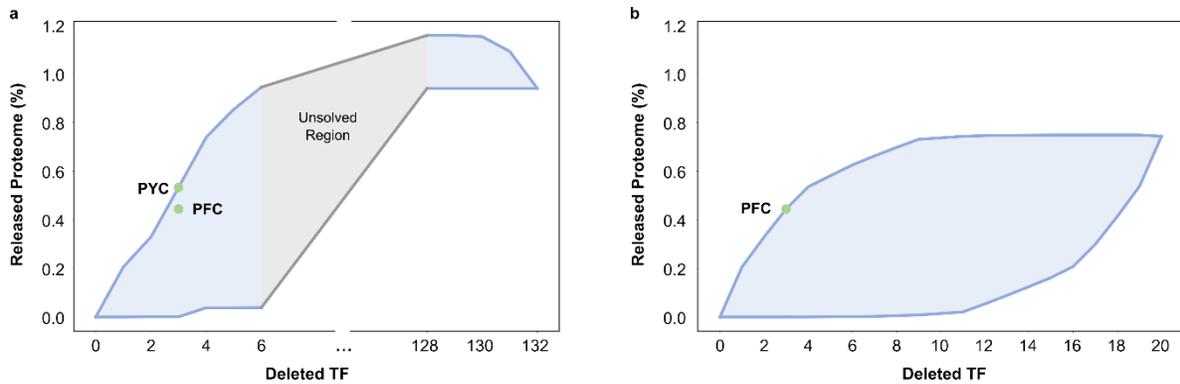


Figura 4 | Cálculos de liberación del proteoma usando ReProMin. **a**, Espacio de liberación de proteoma correspondiente al caso ST; la región resuelta se muestra en azul mientras que la región no resuelta se muestra en gris. **b**, Espacio de liberación de proteoma resuelto para el caso UT. La localización en el espacio de las mutantes generadas se muestran con un círculo verde.

Con el objetivo de probar la flexibilidad de ReProMin, se realizaron cálculos para otras condiciones de crecimiento en las que también se dispone de datos proteómicos, esto incluye el crecimiento en medio mínimo suplementado con galactosa, acetato o glicerol + cas aminoácidos (casAA) y medio rico LB. En cada caso, se utilizó el conjunto de genes esenciales core previamente definido además de los genes específicos de cada medio ambiente, los cuales fueron determinados realizando simulaciones de esencialidad a escala genómica con un modelo metabólico en la condición de crecimiento correspondiente (ver métodos).

Como resultado, se encontraron 164 TF no esenciales para galactosa, 166 para acetato, 171 para glicerol + casAA y 165 para medio rico LB. También se encontró que la mayoría de los TFs identificados (89%) son compartidos entre todas las condiciones probadas. Se identificaron 22 TFs candidatos para galactosa y acetato que pertenecen al caso de UT y tienen un PB positivo. Los cálculos de liberación del proteoma se hicieron usando estos subconjuntos de TFs candidatos, las predicciones muestran que podemos liberar el 0.88% y el 0.81% del proteoma total en galactosa y acetato, respectivamente, con la eliminación de todos estos TFs (Figura 5a, b).

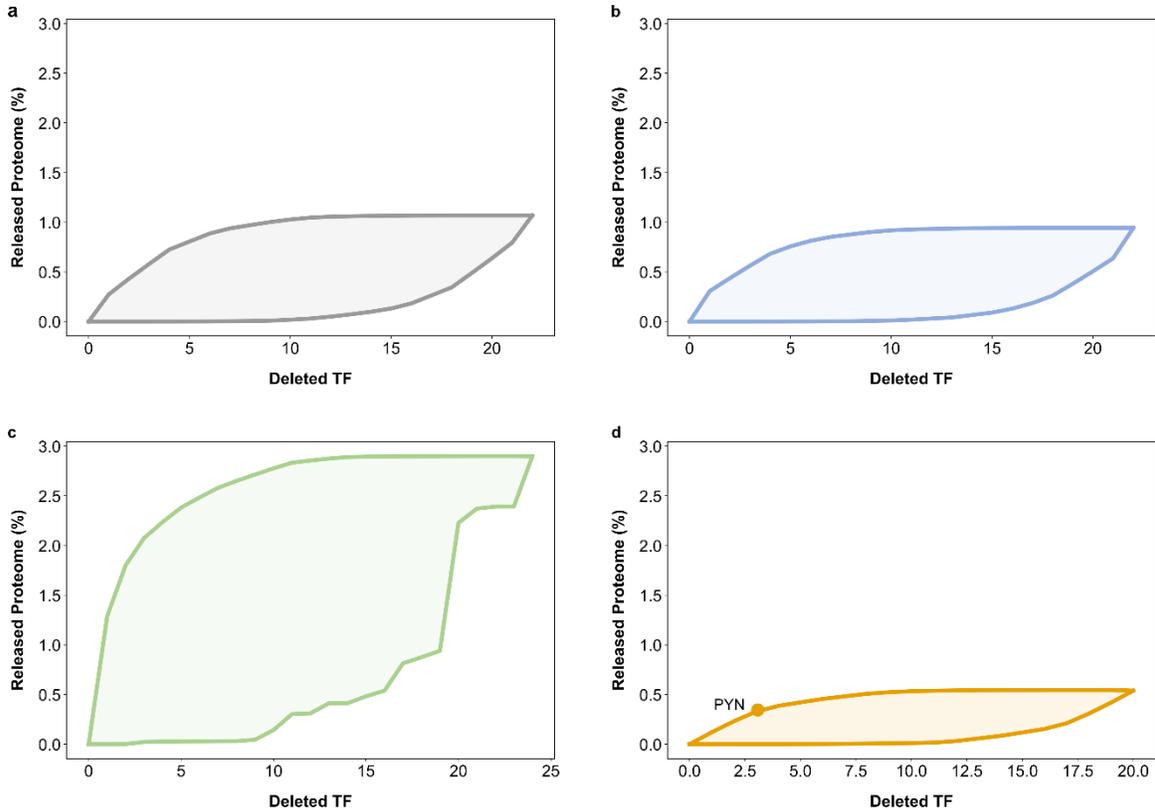


Figura 5 | Espacios de liberación del proteoma calculados por ReProMin correspondientes al caso de UT en distintas condiciones de crecimiento. Espacio de liberación de proteoma correspondiente a **a**, Galactosa, **b**, Acetato, **c**, Glicerol + casAA y **d**, Medio Rico (LB).

Para glicerol, se encontraron 24 TF candidatos, liberando hasta el 2.9% del proteoma total (Figura 5c) y finalmente para medio rico LB fueron identificados 20 TF candidatos, con un potencial de 0.5% de liberación, siendo la peor condición para la liberación potencial del proteoma probado (Figura 5d). Esto concuerda con algunos análisis previos sobre el uso del proteoma, donde se ha demostrado que a mayores tasas de crecimiento hay menos proteoma prescindible^{10,27}.

3.6 Construcción de cepas mutantes knock-out combinatorias.

A pesar de que *E. coli* es uno de los organismos más estudiados y su TRN ha sido ampliamente investigada, sólo la mitad de sus genes tienen información regulatoria depositada en RegulonDB. Con el fin de evitar efectos colaterales perjudiciales debido al conocimiento incompleto de la TRN, se optó por seleccionar la combinación más pequeña de TFs que teóricamente libera la mayor cantidad de recursos celulares. Con base en las predicciones del algoritmo ReProMin, se construyeron dos cepas mutantes, las cuales corresponden a la mejor combinación

de tres TFs de los casos ST y UT predichos para el crecimiento en medio mínimo con glucosa.

Para el caso ST, lo anterior corresponde a la cepa PYC (Δ phoB - sistema de eliminación de fosfatos, Δ yedW - gen de función desconocida, Δ cusR - regulador del sistema de exportación de cobre/plata) con un PB de 1.3 fg que representa el 0.53% del proteoma total en la condición de crecimiento previamente descrita (Figura 4a). Este diseño representa un caso particular de regulación compartida en el que la mayoría de los genes blanco sólo se silencian por la eliminación de los tres TFs en conjunto (Figura 6a).

En el caso UT, la cepa resultante fue nombrada PFC (Δ phoB - sistema de eliminación de fosfatos, Δ flhC - regulador maestro del flagelo, Δ cueR - sistema de eflujo de cobre). Los cálculos muestran que esta cepa posee un PB de 1.08 fg, lo que que representa el 0.44% del proteoma total en glucosa (Figura 4b). El caso UT tiene un mayor grado de confianza en el diseño en comparación con el caso ST debido a una subred reguladora más simple; los genes blanco en este caso pertenecen principalmente al motivo de red de Módulo de Entrada Única (SIM)²⁶ (Figura 6b).

y los círculos grises genes sin cobertura proteómica; el tamaño de los círculos es proporcional al PL del objetivo.

Además, utilizando un enfoque intuitivo, se generó una cepa control en la que fueron eliminados tres TF que también regulan funciones no relacionadas con el crecimiento. Para construir esta cepa se seleccionaron aleatoriamente tres de los TFs que previamente se encontraron regulados a la baja en cepas con mutaciones regulatorias en la RNA polimerasa seleccionadas por Evolución (ALE)¹⁰.

La cepa resultante fue bautizada como FOG (Δ (fliA, oxyR, gadE)) cuyos genes eliminados codifican para el factor sigma del flagelo, el regulador maestro de estrés oxidativo y el regulador de la resistencia a ácidos respectivamente. La cepa FOG no fue generada empleando ReProMin; por lo tanto, las intervenciones reguladoras afectan algunas funciones esenciales (nueve genes esenciales) que participan principalmente en la síntesis *de novo* de los grupos AMP y hemo. La cepa se utilizó como control para la comparación con las cepas diseñadas computacionalmente.

Finalmente, para probar ReProMin en un entorno diferente para el que también se dispone de datos de alta confianza sobre esencialidad de genes, se construyó la mejor cepa 3KO en medio rico LB (Figura 5d). La cepa resultante fue PYN (Δ phoB - sistema de eliminación de fosfatos, Δ yqhC - reducción de furfural²⁸ y Δ ntrC - sistema de regulación de nitrógeno de dos componentes) liberando potencialmente el 0.33% del proteoma total.

3.6 El análisis de RNA-seq confirmó la alta especificidad de las mutaciones introducidas.

El poder de predicción de ReProMin depende de la exactitud de las interacciones capturadas en la TRN de *E. coli*. Se validaron los cambios transcripcionales predichos para el caso PFC que mostró la mayor liberación del proteoma, para esto se midieron los perfiles de transcripción de la cepa silvestre y la cepa PFC obtenidos por el RNA-seq en el mismo ambiente. El objetivo de este experimento es determinar el grado de éxito en el silenciamiento de los genes a nivel transcripcional y evaluar otras posibles perturbaciones transcripcionales resultantes de las modificaciones regulatorias.

Los resultados muestran que en la cepa PFC no se detectaron transcritos correspondientes a los tres TFs eliminados (Figura 7a). Al mapear el cambio de expresión obtenido a las predicciones de la herramienta computacional, es posible visualizar el impacto a nivel transcripcional de los reguladores faltantes sobre sus objetivos (Figura 7c).

Cuatro blancos asociados a flhC, correspondientes a los genes que conforman el flagelo (flgB, flgC, flgE y flgG) fueron completamente silenciados; además, el resto de genes relacionados con esta estructura también registraron una disminución en su expresión.

En cuanto phoB, se silenciaron con éxito dos blancos (phnI y phnL), ambos genes pertenecen a un operón que se induce bajo inanición de fosfato y es necesario para el uso de fosfonato y fosfito como fuentes de fósforo²⁹, además muchos otros genes relacionados con este operón también redujeron su expresión. Por el contrario, phnK presente en el mismo operón sorprendentemente se encontró sobre expresado. Finalmente, no se encontraron transcritos de seis genes pertenecientes al operón mencionado anteriormente, los cuales podrían no estar totalmente expresados en ausencia de inanición de fosfato. Además, phoR (parte del sistema de dos componentes phoB-phoR) también redujo su expresión.

Por último, ambos objetivos de cueR (copA y cueO) también disminuyeron drásticamente su abundancia.

En cuanto a la exactitud de las predicciones de ReProMin, 30 genes de 47 genes predichos fueron silenciados a diferentes niveles, mientras que 9 genes silenciados predichos presentaron valores de expresión más altos que la cepa WT. Al examinar de manera cautelosa las discrepancias entre los cambios transcripcionales predichos y medidos, se observó que algunas interacciones que construyen la red reguladora de *E. coli* fueron predichas computacionalmente o se derivan de observaciones indirectas. Esas evidencias débiles dieron lugar a falsas predicciones, sin embargo, lo anterior puede minimizarse si se mejora la caracterización de la red. Por último, no se encontraron transcritos de 8 objetivos predichos en ninguna de las dos cepas. Estas observaciones muestran que en el 72% de los casos (28 de 39 genes medidos) las predicciones de la herramienta computacional fueron exactas (Figura 7d).

Además de los cambios transcripcionales diseñados, se encontraron 17 genes expresados de forma diferencial (8 genes regulados hacia abajo y 9 regulados hacia arriba) (Figura 7b; Tabla 4). Este análisis de RNA-seq muestra que, además de los cambios transcripcionales previstos, se identificaron pocos efectos colaterales fuera del objetivo a nivel transcriptómico en la cepa PFC durante el crecimiento en medio M9 con glucosa.

Figura 7 | Análisis transcriptómico de la cepa diseñada para el caso UT. **a**, Gráfica de correlación de los transcritos para las cepas PFC y WT. Los cuadrados verdes representan los tres TFs eliminados. **b**, Volcano plot que muestra la expresión diferencial de los genes. En ambos casos, se destacan los genes estadísticamente significativos (azul - regulado hacia abajo, amarillo - regulado hacia arriba) (\log_2 Fold Change ≥ 1 o ≤ -1 , $P \leq 0,01$, $n = 2$). **c**, Integración de los datos de transcriptómica con las predicciones computacionales de ReProMin. El tamaño del círculo corresponde al cambio en la expresión de cada objetivo (los círculos más grandes representan genes totalmente silenciados), en todos los casos los círculos azules representan objetivos que liberan recursos (regulados hacia abajo), los círculos amarillos representan objetivos que generan carga proteómica (regulados hacia arriba) y los círculos grises objetivos que no se encontraron expresados. **d**, Exactitud de las predicciones de ReProMin basadas en datos de RNAseq. Los círculos amarillos representan predicciones erróneas, los círculos azules representan predicciones precisas y los círculos grises representan predicciones no mapeadas (la expresión no se detectó).

3.7 Cuantificación de la masa del proteoma liberado.

Para cuantificar de manera indirecta la masa real de la fracción de proteoma liberada a partir de datos de RNA-seq, se realizó una estimación de las tasas de traducción de los blancos asociados a los TFs eliminados además de los genes expresados diferencialmente previamente discutidos (Figura 8a, ver métodos).

Los cálculos mostraron que la masa de la fracción del proteoma asociada a los blancos de los TFs eliminados disminuyó de un 0.48% del proteoma total en la cepa WT a un 0.25% en la cepa PFC, siendo CueR el TF que más contribuye a la liberación (Figura 8b). Inicialmente los cálculos de ReProMin estimaron una liberación del 0.44% del proteoma total en PFC, según estos cálculos y considerando la incertidumbre que puedan tener, correspondería a un 57% de éxito en la liberación de recursos previamente estimada. Los recursos liberados aparentemente se redistribuyen a los genes que se encontraron sobre expresados, los cuales muestran una mayor fracción del proteoma en la cepa mutante (Figura 8c); estos recursos podrían quedar disponibles para la expresión de funciones heterólogas.

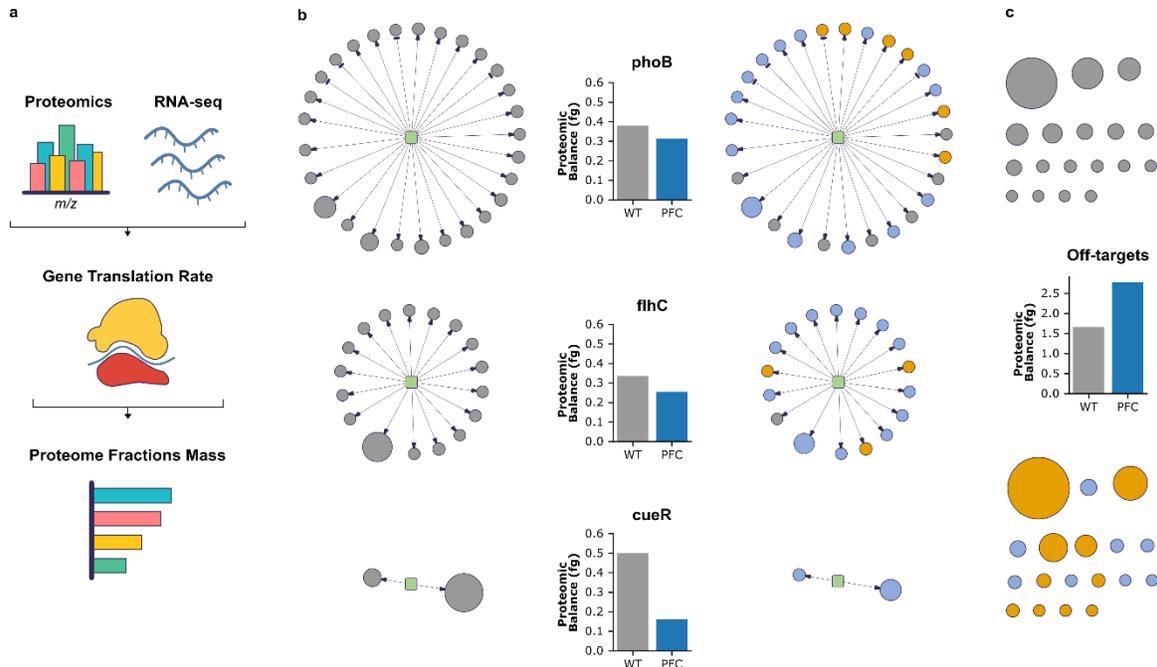


Figura 8 | Estimación de la masa de la fracción del proteoma asociada a los TFs eliminados. **a**, Metodología utilizada para estimar los cambios del proteoma a partir de los datos de RNA-seq. **b**, Representación de la masa de la fracción del proteoma y comparación de los objetivos asociados a los TFs eliminados para la cepa WT (izquierda) y PFC (derecha). **c**, Representación de la masa de la fracción del proteoma y comparación de los objetivos identificados para la cepa WT (arriba) y PFC (abajo). Los cuadrados verdes representan los TF suprimidos; en el caso del PFC, los círculos azules significan objetivos que redujeron su masa en comparación con el WT, los círculos amarillos objetivos que aumentaron su masa y los círculos grises objetivos no encontrados expresados, por lo que no contribuyen a la masa; el tamaño de los círculos es proporcional al PL del objetivo. En todos los casos las barras del gráfico representan la suma de la masa de todos los genes de cada grupo.

3.8 Evaluación fenotípica de las cepas diseñadas.

Las tres cepas mutantes generadas experimentalmente (caso ST, caso UT y control) fueron evaluadas en medio rico (LB) así como en medio mínimo (M9) adicionado con una de tres fuentes de carbono diferentes (acetato, galactosa, glucosa).

Las cepas diseñadas por ReProMin (PYC y PFC) no mostraron defectos de crecimiento ni aumento de la tasa máxima de crecimiento o del rendimiento de la biomasa en ninguna de las cuatro condiciones probadas. Por el contrario, la cepa de control (FOG) mostró defectos de crecimiento en todas las condiciones de crecimiento probadas (Figura 9). Este defecto fenotípico mostrado por la cepa FOG

puede ser el resultado de la eliminación del principal activador del sistema de resistencia a los ácidos (*gadE*), ya que se empleó crecimiento en lote sin control de pH.

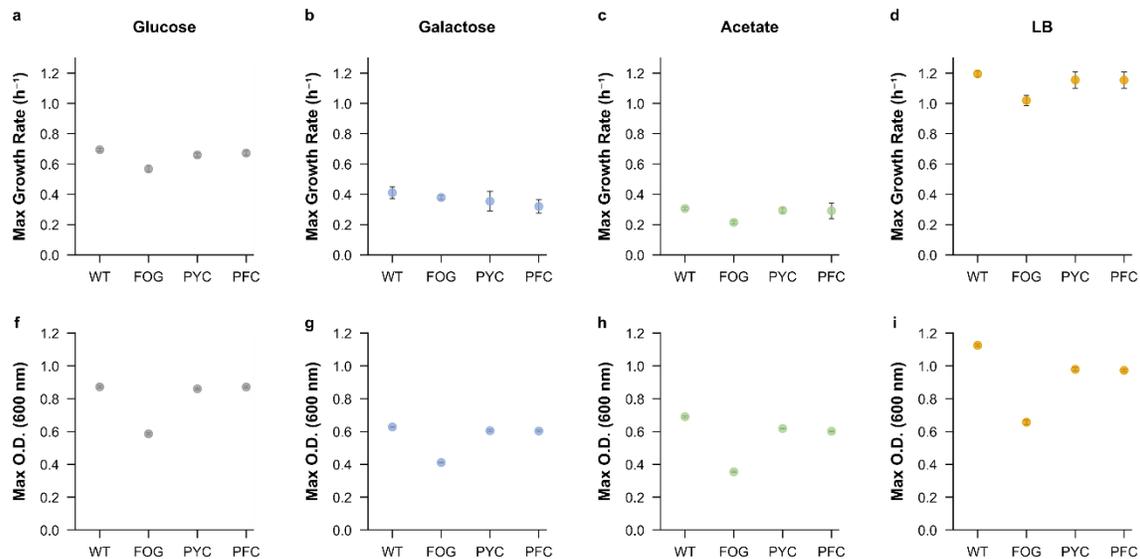


Figura 9 | Evaluación fenotípica de las cepas basada en las predicciones de ReProMin para glucosa (casos de UT y ST) y el control. Crecimiento en medio M9 suplementado con diferentes fuentes de carbono y medio rico (LB). **a-d**, muestra la tasa máxima de crecimiento y **e-h**, muestra la densidad óptica (O.D.) máxima. Los puntos representan el valor gaussiano ajustado ± 2 s.d. para $n = 9$.

En el caso de medio mínimo con glucosa, también se evaluó el efecto sobre el crecimiento durante la producción de proteínas recombinantes, para esto se utilizó un plásmido que expresa un circuito genético con dos reporteros fluorescentes (Figura 11a)³⁰. La carga causada por llevar un plásmido se reflejó como una disminución en la tasa de crecimiento en todas las cepas probadas (Figura 10a); esta disminución es mayor cuando el plásmido expresa el circuito genético, sin embargo, la carga mostrada por ambas cepas diseñadas por ReProMin fue menor en comparación con la contraparte de tipo silvestre. Además, la cepa PFC también mostró una mayor producción de biomasa final (Figura 10b).

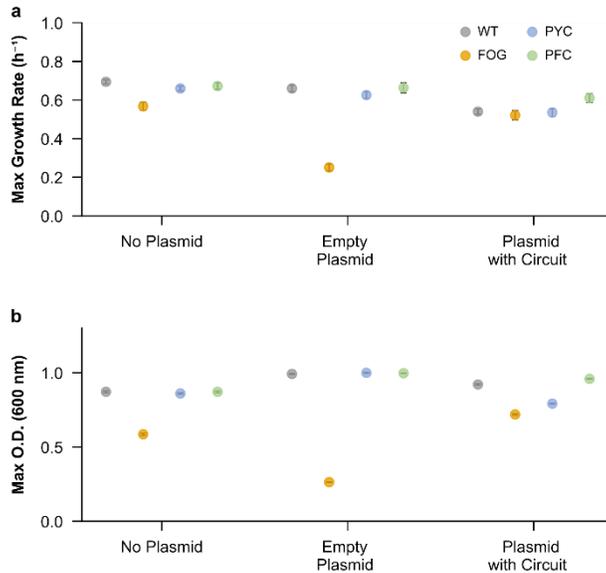


Figura 10 | Evaluación de la carga metabólica de las cepas basadas en las predicciones de ReProMin para medio mínimo con glucosa (casos UT y ST) y el control. Efecto en el crecimiento mientras se porta un plásmido vacío y un plásmido expresando un circuito sintético, **a**, muestra la tasa máxima de crecimiento y **b**, muestra la densidad óptica (O.D.) máxima. Los puntos representan el valor gaussiano ajustado \pm 2 s.d. para $n = 9$.

Se ha descrito que los niveles de expresión de dos proteínas reporteras codificadas en el mismo plásmido, pero sin una conexión regulatoria entre ellas, es capturado por una relación lineal que puede ser interpretada como una línea de isocosto, un concepto usado en microeconomía para describir cómo dos productos pueden ser comprados con un presupuesto limitado, de manera que cuanto más se invierte en un producto, menos se puede invertir en el otro. Estas líneas representan el límite del presupuesto de producción de una determinada cepa y condición de crecimiento (Figura 11b)³⁰.

Se obtuvieron las líneas de isocosto en crecimiento balanceado, definido como el periodo durante la fase de crecimiento exponencial en la que se logra un estado constante de la concentración celular de ambos reporteros (en otras palabras, la concentración celular de ambas proteínas fluorescentes no cambia en el tiempo, ~después de 5 horas); estas líneas fueron determinadas por dos métodos diferentes: la fluorescencia media medida por medio de un lector de placas y la fluorescencia media medida por citometría de flujo.

En todos los casos, la línea correspondiente a la cepa PFC muestra un desplazamiento paralelo ascendente en comparación con la cepa WT, lo que representa un aumento del 9% en la fluorescencia absoluta (Figura 11c), del 5% en

la fluorescencia normalizada por célula (Figura 11d) y del 12% en la fluorescencia media por célula (Figura 11e). La diferencia se hace más notoria en la fase estacionaria del cultivo (~22 hrs) donde se alcanza una biomasa máxima más alta y la cantidad de proteína recombinante se incrementa hasta el 18% (Figura 13a).

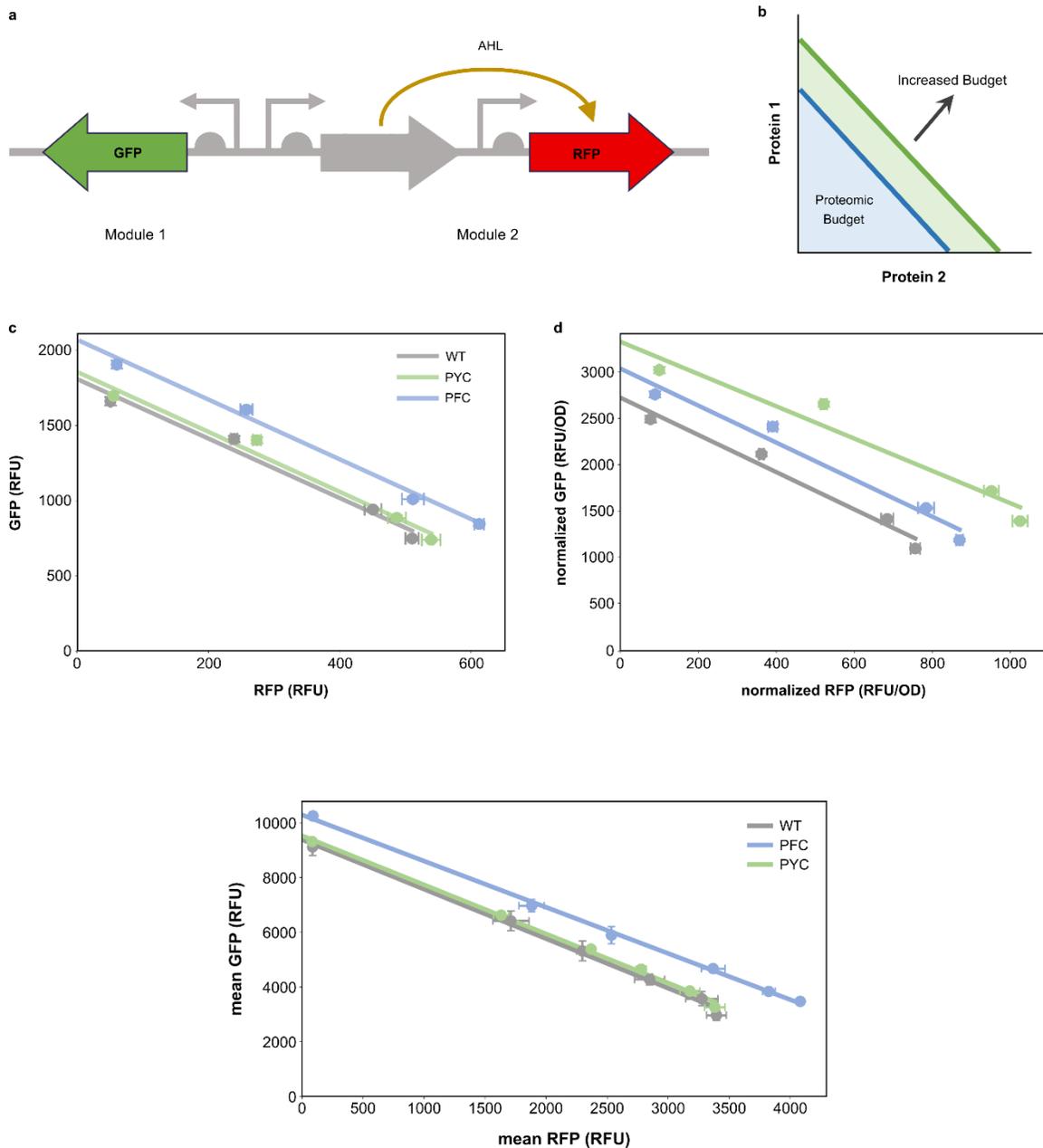


Figura 11 | Caracterización del circuito sintético. **a**, Esquema del circuito genético evaluado que codifica los reporteros fluorescentes GFP y RFP: el primero se expresa constitutivamente, mientras que el segundo está bajo el control de un promotor inducible por N-acil homoserin lactona (AHL). **b**, Al graficar la expresión de una proteína contra la

otra a diferentes niveles de inducción, se obtiene una línea de isocosto. El tamaño del área debajo de la línea representa el presupuesto total del proteoma dedicado al circuito, un desplazamiento hacia arriba de la línea representa un aumento del presupuesto. Líneas de isocosto durante la fase de crecimiento balanceado (~5 h) de las cepas diseñadas mostrando **c**, fluorescencia absoluta, **d**, fluorescencia normalizada y **e**, fluorescencia media por célula medida por citometría de flujo. En todos los casos, los puntos representan el valor gaussiano de fluorescencia ajustado ± 2 s.d. para $n = 9$ del reportero rojo (eje x) graficado contra el reportero verde (eje y) en una concentración creciente de inductor (0, 2.5, 5, 20 nM AHL). Se utilizó una regresión lineal para ajustar los puntos a una línea.

La cepa del caso ST (PYC) no mostró ningún aumento de presupuesto comparado con la cepa WT en fluorescencia absoluta (Figura 11c), sin embargo, mostró un aumento de presupuesto en fluorescencia normalizada (Figura 11d). La mutante PYC muestra una reducción de la producción máxima de biomasa (Figura 11e), estos resultados sugieren que hay un aumento del presupuesto para producir proteína recombinante, pero también una reducción de la capacidad de crecimiento probablemente debido a las inesperadas interacciones regulatorias más complejas del caso ST discutidas previamente.

Las simulaciones con el modelo ME previamente mostraron que al reducir la fracción de proteína no utilizada (UPF) de 0.36 a 0.25, se aumenta el sector máximo del proteoma recombinante de 0.15 a 0.20 (Figura 1). Este aumento en la producción de proteína recombinante concuerda con observaciones experimentales de trabajos anteriores³¹, en los que para la condición de crecimiento en medio mínimo con glucosa la fracción heteróloga máxima observada fue del 14% del proteoma total, por lo que un aumento del 0.5-1% en la disponibilidad del proteoma puede dar lugar a un aumento del 10% en el presupuesto heterólogo, que se acerca a lo que se encontró en la cepa mutante PFC.

También se construyó la mejor combinación de 4KO para medio mínimo con glucosa. ReProMin predice PFC con la adición de marA como el mejor 4º TF a eliminar, con un 0.54% de PB. La cepa resultante: PFCM (Δ marA: resistencia a múltiples antibióticos) también tiene un mayor presupuesto proteómico que la cepa silvestre, sin embargo, no es mayor que el de la cepa PFC (Figura. 12a), confirmando la hipótesis de que un menor número de intervenciones genéticas tendrá una mayor probabilidad de éxito debido a las interacciones epistáticas de las múltiples eliminaciones de TF.

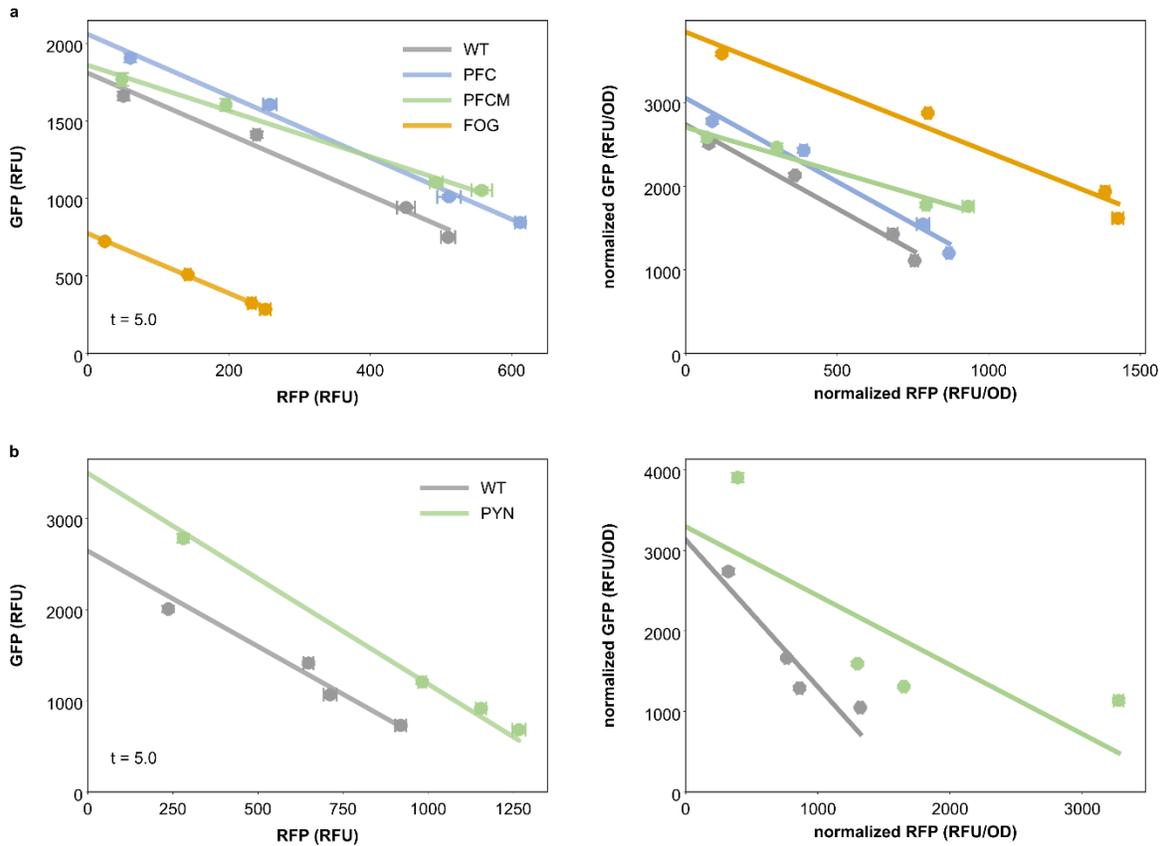


Figura 12 | Líneas de isocosto durante el crecimiento equilibrado (~5 h). Líneas de isocosto de las cepas mutantes generadas para dos condiciones de crecimiento: **a**, medio M9 con glucosa M9 y **b**, medio rico (LB). Izquierda: fluorescencia absoluta, Derecha: fluorescencia normalizada

Además, para probar experimentalmente las predicciones de ReProMin en un entorno diferente, se evaluó la mutante PYN descrita anteriormente en medio rico LB. PYN tiene un mayor presupuesto de proteoma que la cepa de tipo silvestre durante el crecimiento equilibrado en LB, y también en la fase estacionaria (Figura 12b, 13b) lo que demuestra que ReProMin puede ser utilizado en cualquier condición donde la esencialidad de los genes y la información proteómica estén disponibles.

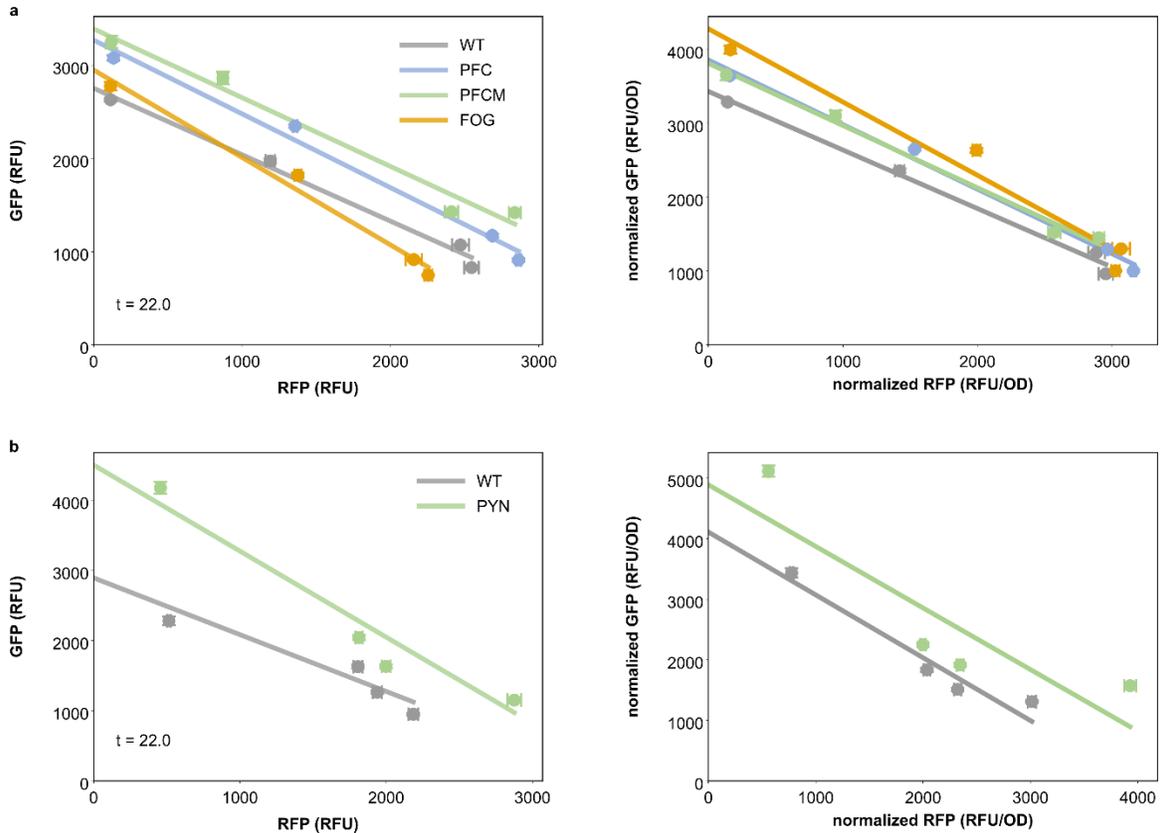


Figura 13 | Líneas de isocosto durante fase estacionaria (~22 h). Líneas de isocosto de las cepas mutantes generadas para dos condiciones de crecimiento: **a**, medio M9 con glucosa M9 y **b**, medio rico (LB). Izquierda: fluorescencia absoluta, Derecha: fluorescencia normalizada

3.9 Expresión de una vía metabólica sintética: producción de violaceína.

Como prueba de concepto para posibles aplicaciones biotecnológicas, se evaluó la capacidad de la cepa PFC para sintetizar la molécula violaceína a partir de una costosa vía metabólica heteróloga. La violaceína es un pigmento sintetizado por *Chromobacterium violaceum* y está dotado de numerosas actividades biológicas (antibacteriana, antiviral, antiparasitaria) y ha cobrado recientemente importancia en el ámbito industrial, especialmente para aplicaciones en cosméticos, medicamentos y textiles³².

La violaceína se sintetiza en una vía metabólica de cinco pasos utilizando triptófano como precursor. En este caso, se utilizó el plásmido conteniendo la vía de la violaceína ensamblado por Darlington *et al.*³³, donde los cinco genes para la biosíntesis de esta molécula están dispuestos en dos operones, uno de los cuales consiste en *vioA* expresado constitutivamente, mientras que el resto de la vía codificada por los genes *vioBCDE* está bajo el control de un promotor inducible por

AHL (Figura 14a). Esta construcción sigue el mismo principio que el circuito anterior, de modo que cuanto más se produce un módulo, menos se expresa el otro debido a la competencia por los limitados recursos para la expresión de los genes. Sin embargo, en este caso el número de genes en cada módulo es diferente y codifica para enzimas metabólicamente activas con diferentes propiedades cinéticas, lo que da lugar a una biosíntesis diferencial de violaceína. Este sistema es ideal para probar las cepas diseñadas por ReProMin, ya que su capacidad de producir un metabolito a partir de una vía heteróloga depende directamente de la fracción de la maquinaria celular que puede asignarse para la expresión del proteoma accesorio en una serie de condiciones (de baja a alta competencia).

Se caracterizó la producción de violaceína después de 24 horas en la cepa silvestre y PFC utilizando medio M9 + glucosa con y sin triptófano (2,0 g/l), en ambos casos se añadió AHL (1,25, 2,5, 5, 10, 20 nM) para la inducción. Se encontró que la máxima producción se alcanzó con una cantidad mínima de inductor (1,25 nM), lo que indica que es crucial tener una expresión equilibrada de la vía con la cantidad adecuada de cada módulo para maximizar la síntesis del producto final. La cepa PFC mostró un aumento medio en la producción de mg de violaceína del 18% ($p < 0,05$) (Figura 14b) y del 20% en los datos normalizados (mg de violaceína/mg de proteína) (Figura 15a).

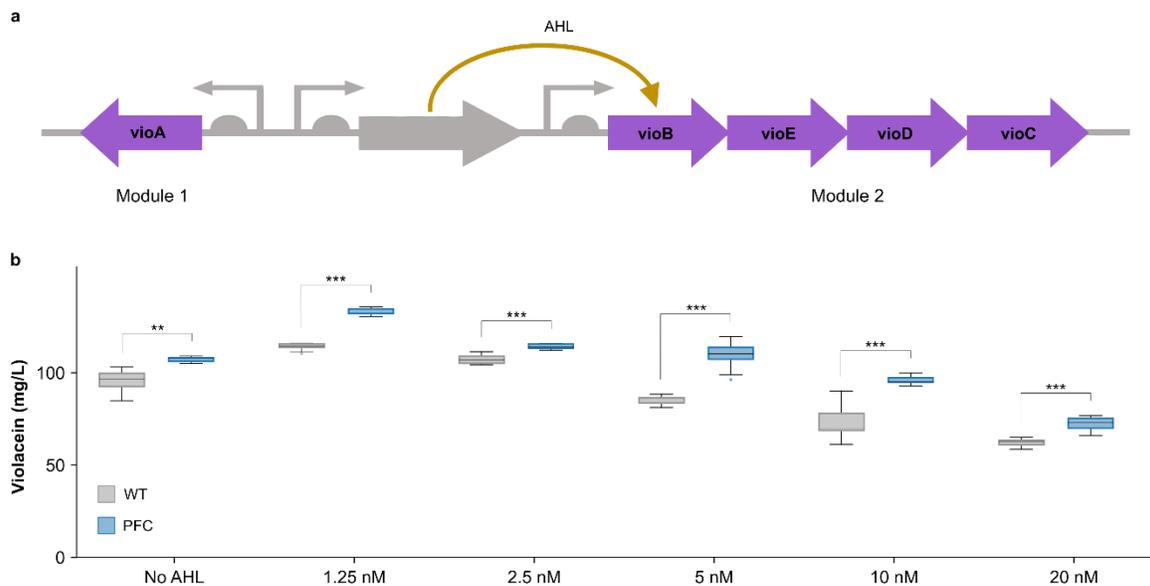


Figura 14 | Evaluación de la vía de la violaceína. a, Esquema del circuito que expresa la vía para la biosíntesis de violaceína. **b,** Producción total de violaceína usando 2 g/L de triptófano después de 24 h en presencia de concentraciones crecientes de inductor (AHL) (media \pm s.d., n = 9). Los asteriscos *, ** y *** denotan diferencias significativas entre WT y PFC usando una prueba t de Student de dos colas sin emparejar. Los siguientes valores de

P se obtuvieron para las concentraciones de AHL: No AHL, $P = 0,0031$; 1.25 nM, $P < 0,0001$; 2.5 nM, $P < 0,0001$; 5 nM, $P < 0,0001$; 10 nM, $P < 0,0001$; 20 nM, $P < 0,0001$.

Aunque las intervenciones reguladoras diseñadas no se dirigen a la red metabólica del organismo, se encontró un aumento significativo en la producción de violaceína, incluso sin la adición de triptófano en un fondo no diseñado metabólicamente y sin una mayor optimización de la condición de cultivo (Figura 15b). Este aumento, presumiblemente resultante de una mejor expresión de la vía heteróloga, muestra que ReProMin también puede aprovecharse para aumentar la producción de metabolitos de costosas vías metabólicas heterólogas.

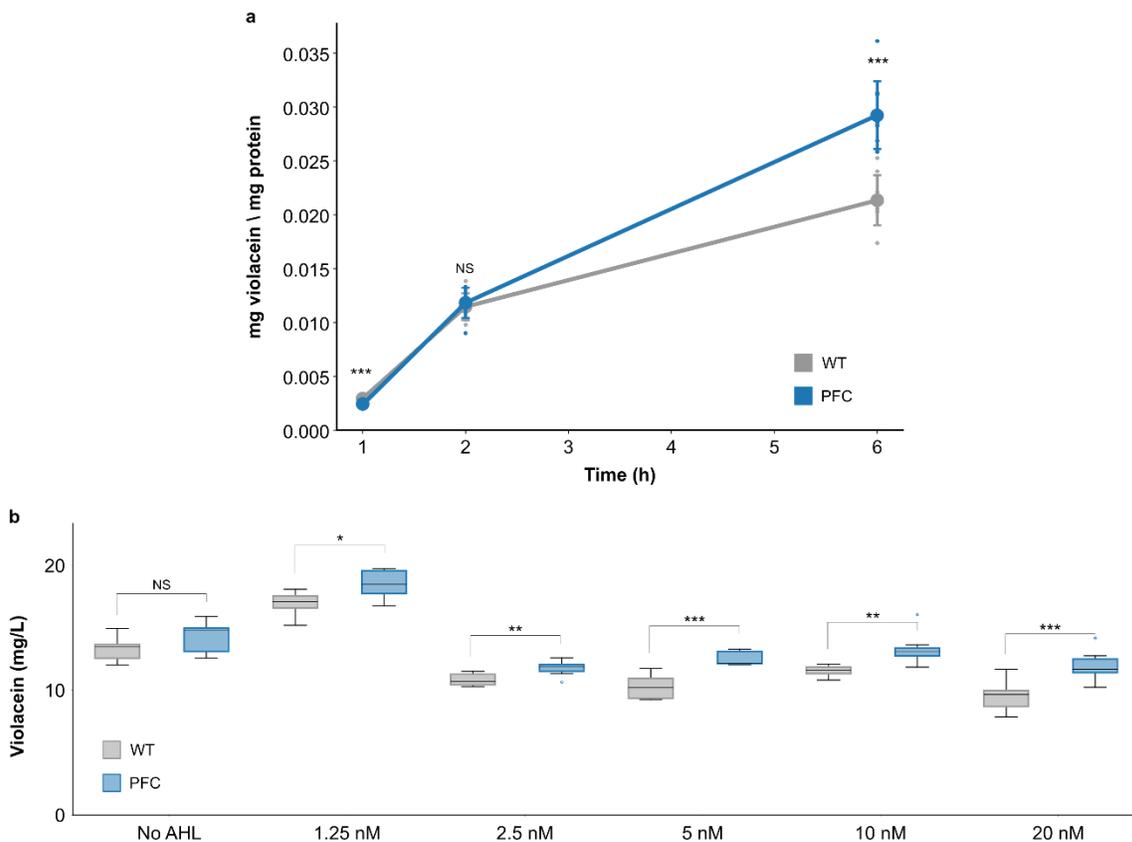


Figura 15 | Caracterización de la producción de violaceína. a, Producción de violaceína normalizada con 2 g/L de triptófano en presencia de AHL (20 nM) (media \pm s.d., $n = 9$). **b,** Producción total de violaceína sin añadir triptófano después de 24 h en presencia de concentraciones crecientes de inductores (AHL) (media \pm s.d., $n = 9$). Los asteriscos *, ** y *** denotan diferencias significativas entre WT y PFC usando una prueba t de Student de dos colas sin emparejar. Los siguientes valores de P se obtuvieron para la producción

normalizada de violaceina: 1 h, $P = 0,0003$; 2 h, $P = 0,5647$; 6 h, $P < 0,0001$. Los siguientes valores de P se obtuvieron para la producción de violaceina con diferentes concentraciones de AHL: No AHL, $P = 0,0599$; 1,25 nM, $P = 0,0146$; 2,5 nM, $P = 0,0021$; 5 nM, $P < 0,0001$; 10 nM, $P = 0,0014$; 20 nM, $P = 0,0005$.

Capítulo 4

Discusión

Las redes de regulación genética son estructuras robustas las cuales pueden modificarse con resultados fenotípicos interesantes³⁴, es por esto que son un objetivo de ingeniería perfecta para aplicaciones de biología sintética, al eliminar los activadores del proteoma de cobertura es posible liberar recursos celulares y aumentar la capacidad y eficiencia de funciones sintéticas.

En este trabajo se ha demostrado que mediante la definición de un conjunto de genes esenciales combinado con información de regulación genética permite la identificación de TFs cuya eliminación conduce directamente al silenciamiento de las fracciones de proteoma que no se usan en una condición de crecimiento en particular, además, mediante la integración de datos proteómicos cuantitativos, es posible calcular la cantidad de recursos liberados por cualquier TF.

La herramienta computacional desarrollada tiene la capacidad de resolver los difíciles casos de regulación cooperativa pasa así encontrar la combinación de TFs que maximizan la liberación de recursos para un numero particular, la visualización de estas combinaciones dentro del espacio de liberación permite evaluar el costo – beneficio de introducir una nueva mutación. En este caso se encontró que la introducción de solo tres mutaciones es lo más óptimo, también se observó que la cantidad total de proteoma para liberar depende de la condición de crecimiento, siendo 1.3% el mejor de los casos analizados.

La cepa diseñada por ReProMin PFC mostró un mayor presupuesto proteómico, medido mediante líneas de isocosto, también se observó una mayor capacidad para producir un metabolito a partir de una ruta heteróloga. Aunque la cantidad de proteoma reducido en la cepa de PFC puede parecer insignificante en las condiciones evaluadas, se observó que una reducción de <1% del proteoma no utilizado puede aumentar la fracción de proteoma heterólogo en un 10%.

El perfil de transcripción obtenido por RNA-seq muestra la exactitud de ReProMin ya que los genes eliminados fueron silenciados por completo además de que se encontraron pocos blancos afectados de manera colateral.

Al comparar las cepas diseñadas por medio de ReProMin con la cepa mutante control (FOG), fue posible mostrar que la eliminación imprecisa de TF produce efectos perjudiciales sobre el crecimiento, la producción máxima de biomasa y proteína heteróloga, lo que indica que la eliminación de TFs no es un proceso trivial ya que puede afectar funciones esenciales e introducir defectos fenotípicos.

Es necesario considerar que el método tiene ciertas limitaciones, para empezar, el conocimiento de la TRN de *E. coli* aún está incompleto, no todos los genes poseen interacciones depositadas en RegulonDB, además la falta de información basada en condición de crecimiento o en etapa de crecimiento, genera incertidumbre sobre el estado real de expresión de algunos genes por lo que se debe suponer que las interacciones regulatorias están presentes en todo momento, debido a esto es que se decidió profundizar más en el estudio del caso UT sobre el ST. Por otro lado, no todos los genes tienen información proteómica por lo que el cálculo del balance proteómico a veces no es exacto, no es posible de determinar o está incompleto. La certeza de ReProMin depende de la calidad y la especificidad de los datos usados en el análisis.

Se espera que los avances en tecnologías de alto rendimiento (como Chip-seq) combinados con nuevos enfoques computacionales³⁵⁻³⁷ permitirán la rápida generación de redes reguladoras completas, que combinadas con datos proteómicos de alto rendimiento o tasas de traducción de genes obtenidas por Ribo-Seq, mejorarán la precisión del método y permitirán su aplicación incluso en organismos no modelo.

De acuerdo con la lista de genes esenciales y comparando las predicciones de uso del proteoma del modelo ME con los datos proteómicos reales, se calcula que un 33.6% del proteoma puede ser eliminado directamente, esto concuerda con las observaciones de O'Brien *et al.*, quienes demostraron que en medio mínimo con glucosa hasta el 30% del proteoma no es utilizado²⁰. Desafortunadamente, la mayoría de los genes que contribuyen a la fracción del proteoma no esencial no tienen información regulatoria conocida o están regulados por un TF clasificado como esencial, por lo tanto, están fuera del alcance de ReProMin. Reducir el 30% completo del proteoma de cobertura sigue siendo un desafío que puede abordarse mediante una combinación de enfoques, como la eliminación de reguladores globales de la transcripción, la eliminación directa de genes altamente expresados o la optimización del proteoma core³⁸.

Previamente, se han abordado varios enfoques experimentales para la optimización de la asignación de recursos en bacterias. La minimización del genoma se ha realizado principalmente mediante intervenciones genéticas a gran escala cuyos resultados son difíciles de predecir y no muestran una mayor estabilidad del genoma^{39,40}, por ejemplo, al analizar los datos de una cepa de *E. coli* con genoma reducido⁴¹, se encontró que una reducción del genoma del 15% (743 genes) condujo a una reducción del proteoma (de acuerdo con cálculos realizados con el mismo conjunto de datos proteómicos) de sólo el 1.5% en medio mínimo con glucosa, ReProMin logro liberar un tercio de ello con solo tres mutaciones.

La metodología presentada en este trabajo es una estrategia novedosa para la optimización del proteoma con intervenciones genéticas mínimas que supera las serias limitaciones de eliminar grandes regiones del genoma; es una metodología flexible que se puede aplicar a otras condiciones de crecimiento y producción y también a diferentes organismos para los cuales exista suficiente información disponible. Este trabajo muestra el potencial del diseño racional de los sistemas biológicos sobre los enfoques de prueba y error predominantemente utilizados en biología sintética.

Capítulo 5

Perspectivas

A diferencia del método ReProMin en el que se eliminaron los factores de transcripción que regulan genes no esenciales para lograr la reasignación de recursos, en este caso se propone eliminar genes que codifican proteínas funcionales altamente expresadas. Los cálculos del proteoma no utilizado en *E. coli* han mostrado que hasta un 33% del proteoma puede ser eliminado en condiciones de medio mínimo de glucosa⁴².

Para determinar los genes a eliminar, se realizaron simulaciones en un modelo ME, la lista de genes NN está compuesta por todas aquellas proteínas que se consideran no esenciales (N) (según la lista de genes esenciales previamente definida) y no utilizadas (N) por el modelo ME. Al calcular la carga proteómica de las proteínas comprendidas en la lista NN se encontró que la proporción de la carga proteómica entre los genes de la lista es desigual. Se observó que aproximadamente el 80% del total del proteoma no utilizado puede ser liberado silenciando la expresión de los diez primeros genes de la lista NN.

La lista de genes NN es muy similar a los genes encontrados a través de Ribo-seq por Price *et al.*⁴³, los autores definen estas proteínas como el proteoma "en espera" el cual tiene un alto costo de recursos celulares sin aportar ningún beneficio fenotípico.

Para probar el efecto de la eliminación de proteínas no esenciales altamente expresadas en un ambiente definido (medio mínimo M9 con glucosa), se construyeron dos mutantes 3KO: OAO ($\Delta ompC$ - porina de membrana externa, $\Delta aceA$ - isocitrato liasa (shunt de glioxilato) y $\Delta ompT$ - proteasa de membrana externa) liberando en teoría una fracción de 2.6% del proteoma y OLA ($\Delta ompF$ - porina de membrana externa, $\Delta livJ$ - transportador de leucina / isoleucina / valina y $\Delta ahpC$ - alquil hidroperóxido reductasa, subunidad C22) liberando una fracción de 3.1%.

Resultados previos muestran que OLA y OAO registran un aumento del 11% y 15% respectivamente en la capacidad para expresar proteína recombinante, sin embargo, OAO es incapaz de metabolizar acetato como fuente de carbono, debido a la ausencia de *aceA*, un defecto fenotípico que afecta el rendimiento del circuito sintético. El gen *aceA* forma parte del operón *aceBAK*, el cual codifica para las dos enzimas y una cinasa implicadas en el ciclo del glioxilato, un by-pass situado en el segundo paso del TCA, que proporciona una estrategia simple y eficiente para

convertir el acetil-CoA en compuestos anapletóricos y gluconeogénicos al crecer con compuestos C₂ como única fuente de carbono. Estos dos pasos permiten a *E. coli* omitir los pasos de descarboxilación oxidativa del ciclo del TCA, conservando así los esqueletos de carbono para la generación de biomasa.

Al comparar la fracción de proteoma total de cada proteína directamente implicada en el by-pass se encontró que aceA está siempre excesivamente sobrerrepresentado en proporción a la otra enzima aceB. Se plantean dos escenarios para una posible optimización. En condiciones donde el ciclo no se está utilizando (ej. Fuentes ricas de carbono) eliminar por completo el resto del ciclo. En las condiciones en las que si se utiliza el by-pass, se propone insertar una enzima aceA catalíticamente más eficiente, así se podrá reducir la fracción de aceA necesaria para la asimilación de sustratos gluconeogénicos. Al eliminar el aceK se tendrá un ciclo no regulado y al afinar la expresión del ICD, la enzima del TCA donde se da la bifurcación, se podrá alcanzar la cantidad requerida de su forma activa para soportar el crecimiento, pero ahorrando recursos al no tenerlo expresado e inactivo.

Capítulo 6

Métodos

1) Simulaciones con modelo ME

Todas las simulaciones se realizaron con el modelo iJL1678b ME²³. Las reacciones de transcripción y traducción correspondientes para la producción de proteína recombinante (GFP) se agregaron manualmente al modelo utilizando métodos estándar. La fracción y el flujo de proteínas no utilizadas a través de la producción de proteínas recombinantes son variables cambiables en el modelo ME que afectan la tasa de crecimiento prevista y la composición del proteoma, los valores de estas dos variables se cambiaron sistemáticamente en el modelo ME para evaluar su efecto sobre la tasa de crecimiento (UPF = 0.36, 0.30, 0.25; Flujo = 0, 0.001, 0.002, 0.0025, 0.0030, 0.0035, 0.0040), todos los demás parámetros del modelo se configuraron como predeterminados. Los sectores de proteoma se clasificaron según O'Brien *et al.*¹²

2) Definición de la lista de genes esenciales

Para compilar la lista de genes esenciales en la condición de medio mínimo de glucosa, se combinaron cinco conjuntos de datos de diferentes fuentes. Tres de ellos se generaron experimentalmente usando diferentes métodos de disrupción génica: A) mutagénesis de transposón usando M9 con glucosa como condición de crecimiento (Tn-seq)⁴⁴, B) eliminando fragmentos grandes del cromosoma usando un sistema de recombinación homóloga en medio rico (LB)⁴⁵ y C) la lista actualizada de los mutantes de la colección Keio que son letales, la colección se generó utilizando medio rico^{46,47}. Se generaron dos listas de genes *in silico* usando simulaciones de modelos metabólicos y de expresión a escala del genoma capaces de predecir las necesidades de expresión génica en una condición particular: D) genes que son esenciales para el crecimiento en M9 con glucosa usando el modelo iOL1554-ME¹² y E) genes que son esenciales para el crecimiento en el modelo metabólico iJO1366 y también experimentalmente en M9 con glucosa¹¹. Dentro de la lista compilada, los genes que pertenecen exclusivamente a las listas de genes de Tn-seq y de simulaciones del modelo ME en medio mínimo con glucosa se consideraron condicionalmente esenciales, ya que estas listas de genes se generaron originalmente usando M9 con glucosa como condición de crecimiento, mientras que el resto de los genes fueron clasificados como esenciales.

Para los casos de galactosa, acetato y glicerol + casAA en condiciones de medio mínimo, se realizaron análisis de genes esenciales con el modelo metabólico iML151543 en COBRApy⁴⁸. Para el rico caso medio LB, se utilizó la lista de mutantes Keio.

3) Identificación de reguladores candidato y análisis combinatorio

Se obtuvieron las interacciones TF-gen depositadas en RegulonDB (versión 8, regulondb.ccg.unam.mx), descartando todas las interacciones factor sigma-gen. A continuación, se clasificaron como esenciales todos los TF que activan al menos un gen esencial (de la lista de genes esenciales específicos de la condición) y como no esenciales todos los TF que no activan ningún gen esencial. Luego se analizó la subred de interacciones de cada TF no esencial analizando numéricamente el grado de salida de cada TF (TF_{OUT}), que se clasifica en salida positiva y negativa (TF_{OUT+} y TF_{OUT-}) que representan genes regulados positivos y negativos respectivamente y el grado de entrada de cada gen regulado ($GENE_{IN}$) a su vez también se divide en positivo y negativo ($GENE_{IN+}$ y $GENE_{IN-}$). Se definió como candidato para la reducción del proteoma todos aquellos TF que activan al menos un gen único, que cumple numéricamente la siguiente condición:

$$TF_{OUT+} \geq 1 \quad \wedge \quad GENE_{n_{IN+}} = 1 \quad \wedge \quad GENE_{n_{IN-}} \geq 0 \quad \text{Ec.1}$$

Donde $GENE_n$, representa cualquier gen activado por el TF.

Por otro lado, el conjunto de datos proteómicos descrito anteriormente¹⁷ se usó para calcular la carga proteómica (PL) de cada gen y el balance proteómico (PB) de cada TF de acuerdo con las ecuaciones de la figura 3.

El análisis combinatorio de ReProMin se llevó a cabo de la siguiente manera, dada una lista de TF, el número total de combinaciones se calculó con esta fórmula:

$$P(n, r) = \frac{n!}{r!(n-r)!} \quad \text{Ec. 2}$$

Donde n representa el número de TF candidatos y r el tamaño de la combinación ($r \leq n$).

Luego, el número total de genes silenciados e inducidos para cada combinación se determinó siguiendo los siguientes criterios: para cada gen involucrado en la

combinación, se restó uno del valor de $GENE_{IN+}$ para cada TF que regula el gen objetivo positivamente y uno para el valor de $GENE_{IN-}$ para cada TF que regula el gen objetivo negativamente. Al final de este proceso, un gen se considera silenciado si:

$$GENE_{IN} = GENE_{IN+} = 0 \quad \text{Ec. 3}$$

O inducido si:

$$GENE_{IN} = GENE_{IN-} = 0 \quad \text{Ec. 4}$$

Finalmente, se calculó y clasificó el PB de cada combinación probada.

El conjunto completo de herramientas computacionales codificadas en Python y los conjuntos de datos utilizados en el análisis están disponibles en el siguiente repositorio (<https://github.com/utrillalab/repromin>). Se usó el software Cytoscape versión 3.7⁴⁹ para trazar la representación de red de los datos.

4) Generación de cepas knock-out combinatorias

Las mutantes combinatorias se generaron por transducción de fago P1 a partir de las cepas de knock-out individuales de la colección Keio de acuerdo con el protocolo descrito por Miller (1992)⁵⁰. La eliminación del casete de resistencia a kanamicina antes de cada transducción se realizó usando el plásmido pE-FLP (plásmido Addgene #45978), pE-FLP fue un regalo de Drew Endy y Keith Shearwinand⁵¹. Cada cepa eliminada se confirmó por PCR usando cebadores que flanquean cada gen. En todos los experimentos, se usó *E. coli* BW25113 como fondo de tipo silvestre. Las características de las cepas, plásmidos y primers utilizados en este estudio se describen en la tabla 5.

5) Extracción y secuenciación de muestras de RNA

Las cepas se cultivaron en 50 ml de medio M9 con glucosa (4 g/l) de medio M9 en cultivos en matraces Erlenmeyer de 250 ml en una incubadora con agitación orbital a 37 °C (250 rpm). Las células se cosecharon en la fase media del crecimiento exponencial usando RNAprotect (Qiagen) de acuerdo con las especificaciones del fabricante. Los sedimentos celulares se incubaron con lisozima, Superaseln y proteasa K durante 10 minutos a 37 °C. El RNA total se aisló y purificó utilizando el kit Quick-RNA (Zymo Research) de acuerdo con las especificaciones del fabricante.

La calidad de todas las muestras se inspeccionó en un chip de RNA-bioanalyzer (Agilent). Comenzando con 10 µg de RNA total de cada muestra, se realizó la eliminación del RNA ribosómico con el kit Ribominus (Invitrogen). Para la construcción de las bibliotecas, se utilizó el kit de preparación de muestras de RNA mTruSeq Stranded de Illumina, siguiendo el protocolo HT. Para la secuenciación se utilizó NextSeq 500 v2, con una configuración de 2 x 75 lecturas paired-end y 10 millones de lecturas por muestra.

Las lecturas se mapearon para hacer referencia al genoma *E. coli* MG1655 (RefSeq: NC_000913.3) usando el alineador Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2>). El análisis diferencial final se realizó utilizando la biblioteca Cufflinks (<http://cole-trapnell-lab.github.io/cufflinks>). Los genes con cambio de $\log_2 \geq 1$ se consideraron regulados hacia arriba y ≤ -1 regulados hacia abajo, considerando un valor $p \leq 0.01$.

6) Estimación del proteoma teórico a partir de datos de RNA-seq

Para hacer una estimación de las tasas de traducción de cada gen a partir de los datos de proteómica y de las abundancias de transcripción, se asumió que los transcritos en los datos de RNA-seq producen el proteoma reportado por Schmidt et al. en glucosa. Se calculó cada tasa de eficiencia de traducción (s_i) utilizando esta ecuación:

$$s_i = \frac{C_{celli}}{r_i}$$

Donde r_i representa el valor de FPKM de cada gen y C_{celli} el número de copias de proteínas reportadas. Para todos los genes sin datos proteómicos, se asumió una tasa fija que corresponde a la tasa promedio de los genes reportados. Luego se utilizaron estas tasas para estimar el número de copias de proteínas por célula (P_i) para la mutante de acuerdo con:

$$P_i = r_i \times s_i$$

Finalmente, se estimó la fracción de masa de proteoma correspondiente a los genes regulados hacia abajo previamente identificados por RNA-seq.

7) Caracterización fenotípica

Para la evaluación del crecimiento en diferentes fuentes de carbono, se usaron las siguientes condiciones: medio M9 glucosa (4 g/L), medio M9 galactosa (3.2 g/L), medio M9 acetato (2.5 g/L) y medio rico LB. Las células se cultivaron durante la

noche en el medio correspondiente. Al día siguiente, las cepas se diluyeron a una OD₆₀₀ de 0.05 en medio fresco y se transfirieron 150 µl del cultivo fresco a una placa transparente de 96 pozos (Corning) y se incubó a 37 °C con agitación lineal rápida en un lector de microplacas (Synergy 2.0, BioTek) durante 24 horas, tomando mediciones para OD₆₀₀ cada 20 min. En todos los ensayos se incluyeron tres réplicas y el experimento se repitió de forma independiente en tres días diferentes.

La caracterización de la cinética de crecimiento se realizó utilizando el algoritmo desarrollado por Swain *et al.*⁵², con los parámetros por defecto. En todos los casos, al comparar los parámetros obtenidos de las cepas mutantes con el tipo silvestre, el valor p se determinó usando una prueba t de Student de dos colas como índice de confiabilidad estadística.

8) Evaluación del circuito sintético

Las cepas se inocularon en medio M9 glucosa con gentamicina (20 ug/ml) y se cultivaron durante la noche. Al día siguiente, los cultivos se diluyeron a una OD₆₀₀ de 0.05 en medio fresco conteniendo AHL (Sigma-Aldrich, St. Louis, MO, EE. UU., concentraciones finales de 1.25, 2.5, 5, 10, 20 nM), luego 150 µl se transfirieron a una placa negra de 96 pozos con fondo transparente (Corning) y se incubó como se describió anteriormente, tomando medidas para OD₆₀₀, GFP (ex., 485 nm, em., 528 nm) y RFP (ex., 590 nm, em., 645 nm). La caracterización de la cinética de producción de GFP y RFP también se realizó utilizando el algoritmo descrito anteriormente.

9) Mediciones de citometría de flujo

Para las mediciones de citometría de flujo, los cultivos celulares se prepararon como se describió anteriormente, pero luego se cultivaron en placas de 24 pozos utilizando 1 ml de medio. Cada hora se tomaron alícuotas de 50 µL de cada pozo y se mezclaron con 150 µL de PBS, el volumen de los pozos se mantuvo constante mediante la adición de medio fresco. La suspensión celular se cargó en un citómetro de flujo Attune NxT (ThermoFisher, Waltham, MA, EE. UU.) y se configuró para determinar GFP (excitación 488 nm; emisión 525/50 nm) y RFP (excitación 561 nm; emisión 620/15 nm). Para cada muestra, se analizaron 20,000 eventos y se estimaron las medias de la población utilizando el software predeterminado del instrumento.

10) Caracterización de la producción de violaceína

Las cepas se inocularon en medio M9 glucosa gentamicina (20 ug/ml) y se cultivaron durante la noche. Al día siguiente, las cepas se diluyeron a una OD₆₀₀ de 0.05 en medio fresco conteniendo AHL (1.25, 2.5, 5, 10, 20 nM) y triptófano (2.0 g/L), luego se transfirieron 150 µl del cultivo fresco a una placa transparente de 96 pozos e incubada como se describe anteriormente.

Después de 24 h, la placa se centrifugó (13,000 × g, 10 min) y se descartó el sobrenadante de cada pozo. La violaceína se extrajo suspendiendo el sedimento en cada pozo en 150 µl de etanol absoluto e incubando la placa a 95 °C durante 10 minutos seguido de centrifugación (13,000 × g, 10 minutos) para eliminar los sedimentos. La violaceína presente en el extracto se determinó espectrofotométricamente a 575 nm en un lector de microplacas (Synergy 2.0, BioTek) y la cuantificación se realizó usando una curva con un patrón de violaceína comprado (Sigma-Aldrich, St. Louis, MO, EE. UU.)

11) Cuantificación de proteína total

El método Biuret se utilizó para la cuantificación de la proteína total. Se tomó 1 ml de cultivo y se centrifugó (13,000 × g, 10 min). Se recogió el sobrenadante y el sedimento se lavó con 0.2 ml de agua, se resuspendió y se centrifugó nuevamente, se descartó el agua. Los sedimentos se volvieron a suspender en 0.1 ml de NaOH 6 M y se incubaron a 95 °C durante 10 minutos para romper las células e hidrolizar las proteínas. Para realizar la reacción de Biuret, se añadieron 0.1 ml de CuSO₄ al 3.2% a las muestras y se incubaron bajo agitación vigorosa durante 5 minutos. A continuación, las muestras se centrifugaron durante 2 minutos y se colocaron 150 µl de sobrenadante en una placa de 96 pocillos. La absorbancia a 555 nm se midió en un lector de placas (Synergy 2.0, BioTek).

Apéndice

Tabla 1 | Lista de genes esenciales definidos para el crecimiento en medio M9 glucosa. Los genes etiquetados como esenciales core siempre son necesarios para el crecimiento, mientras que los esenciales condicionales sólo son necesarios para el crecimiento en el medio de la glucosa M9. La columna de puntuación representa el número de conjuntos de datos en los que se encuentra el gen.

| ID | Name | Classification | Score | | | | |
|-------|------|-----------------------|-------|-------|------|-----------------------|---|
| | | | | b3734 | atpA | Conditional Essential | 1 |
| b0185 | accA | Core Essential | 5 | b3738 | atpB | Conditional Essential | 1 |
| b3255 | accB | Core Essential | 5 | b3732 | atpD | Conditional Essential | 1 |
| b3256 | accC | Core Essential | 5 | b3737 | atpE | Conditional Essential | 1 |
| b2316 | accD | Core Essential | 5 | b3736 | atpF | Conditional Essential | 1 |
| b0115 | aceF | Conditional Essential | 1 | b3733 | atpG | Conditional Essential | 1 |
| b1094 | acpP | Core Essential | 3 | b3735 | atpH | Conditional Essential | 1 |
| b2563 | acpS | Core Essential | 3 | b0177 | bamA | Core Essential | 2 |
| b0474 | adk | Core Essential | 3 | b2595 | bamD | Core Essential | 2 |
| b2697 | alaS | Core Essential | 4 | b3532 | bcsB | Core Essential | 1 |
| b4084 | alsK | Core Essential | 1 | b3973 | birA | Core Essential | 2 |
| b0049 | apaH | Conditional Essential | 1 | b0126 | can | Core Essential | 2 |
| b0062 | araA | Conditional Essential | 1 | b0032 | carA | Conditional Essential | 1 |
| b0063 | araB | Conditional Essential | 1 | b0033 | carB | Conditional Essential | 1 |
| b2818 | argA | Core Essential | 3 | b3056 | cca | Core Essential | 3 |
| b3959 | argB | Core Essential | 3 | b0175 | cdsA | Core Essential | 5 |
| b3958 | argC | Core Essential | 3 | b4224 | chpS | Core Essential | 1 |
| b3359 | argD | Core Essential | 1 | b1870 | cmoA | Core Essential | 1 |
| b3957 | argE | Core Essential | 3 | b1871 | cmoB | Core Essential | 1 |
| b3172 | argG | Core Essential | 3 | b3974 | coaA | Core Essential | 2 |
| b3960 | argH | Core Essential | 3 | b3634 | coaD | Core Essential | 2 |
| b1876 | argS | Core Essential | 5 | b0103 | coaE | Core Essential | 2 |
| b0536 | argU | Core Essential | 1 | N/A | cohE | Core Essential | 2 |
| b3796 | argX | Core Essential | 2 | b3357 | crp | Conditional Essential | 1 |
| b1166 | ariR | Conditional Essential | 1 | b1552 | cspl | Conditional Essential | 1 |
| b0908 | aroA | Core Essential | 3 | b2696 | csrA | Core Essential | 3 |
| b3389 | aroB | Core Essential | 2 | b0733 | cydA | Core Essential | 1 |
| b2329 | aroC | Core Essential | 3 | b0886 | cydC | Core Essential | 1 |
| b1693 | aroD | Core Essential | 2 | N/A | cydX | Conditional Essential | 1 |
| b3390 | aroK | Conditional Essential | 1 | b1275 | cysB | Conditional Essential | 1 |
| N/A | arrS | Conditional Essential | 1 | b2750 | cysC | Core Essential | 2 |
| b3433 | asd | Core Essential | 5 | b2752 | cysD | Core Essential | 1 |
| b0930 | asnS | Core Essential | 5 | b3607 | cysE | Core Essential | 3 |
| b0928 | aspC | Core Essential | 1 | b2762 | cysH | Core Essential | 3 |
| b1866 | aspS | Core Essential | 5 | b2763 | cysI | Core Essential | 2 |

| | | | | | | | |
|-------|------|-----------------------|---|-------|------|-----------------------|---|
| b2764 | cysJ | Core Essential | 3 | b0590 | fepD | Conditional Essential | 1 |
| b2751 | cysN | Core Essential | 2 | b0589 | fepG | Conditional Essential | 1 |
| b4214 | cysQ | Core Essential | 2 | b2610 | ffh | Core Essential | 3 |
| b0526 | cysS | Core Essential | 5 | b0455 | ffs | Core Essential | 2 |
| b1910 | cysT | Core Essential | 2 | b0684 | fldA | Core Essential | 2 |
| b2478 | dapA | Core Essential | 4 | b3288 | fmt | Core Essential | 5 |
| b0031 | dapB | Core Essential | 4 | b4699 | fnrS | Conditional Essential | 1 |
| b0166 | dapD | Core Essential | 4 | b0048 | folA | Core Essential | 3 |
| b2472 | dapE | Core Essential | 4 | b3058 | folB | Conditional Essential | 1 |
| b3809 | dapF | Core Essential | 2 | b2315 | folC | Core Essential | 3 |
| b3287 | def | Core Essential | 5 | b0529 | folD | Core Essential | 4 |
| b3235 | degS | Core Essential | 3 | b2153 | folE | Core Essential | 4 |
| b2511 | der | Core Essential | 3 | b0142 | folK | Core Essential | 2 |
| b3639 | dfp | Core Essential | 3 | b3177 | folP | Conditional Essential | 1 |
| b1570 | dicA | Core Essential | 2 | b0172 | frr | Core Essential | 5 |
| b1569 | dicC | Conditional Essential | 1 | b0094 | ftsA | Core Essential | 3 |
| b3702 | dnaA | Core Essential | 3 | b2748 | ftsB | Core Essential | 3 |
| b4052 | dnaB | Core Essential | 3 | b3463 | ftsE | Core Essential | 3 |
| b4361 | dnaC | Core Essential | 3 | b3178 | ftsH | Core Essential | 2 |
| b0184 | dnaE | Core Essential | 3 | b0084 | ftsI | Core Essential | 3 |
| b3066 | dnaG | Core Essential | 3 | b0890 | ftsK | Core Essential | 2 |
| b0015 | dnaJ | Core Essential | 1 | b0083 | ftsL | Core Essential | 2 |
| b0014 | dnaK | Core Essential | 1 | b3933 | ftsN | Core Essential | 2 |
| b3701 | dnaN | Core Essential | 3 | b0093 | ftsQ | Core Essential | 3 |
| b4362 | dnaT | Core Essential | 2 | b0089 | ftsW | Core Essential | 3 |
| b0470 | dnaX | Core Essential | 3 | b3462 | ftsX | Core Essential | 2 |
| b4049 | dusA | Core Essential | 1 | b3464 | ftsY | Core Essential | 3 |
| b3640 | dut | Core Essential | 3 | b0095 | ftsZ | Core Essential | 3 |
| b0173 | dxr | Core Essential | 5 | b3340 | fusA | Core Essential | 5 |
| b0420 | dxs | Core Essential | 5 | b1779 | gapA | Core Essential | 2 |
| b4147 | efp | Core Essential | 2 | b2093 | gatB | Conditional Essential | 1 |
| b2779 | eno | Core Essential | 5 | b3429 | glgA | Core Essential | 1 |
| b0583 | entD | Core Essential | 1 | b3430 | glgC | Core Essential | 1 |
| b2566 | era | Core Essential | 5 | b3176 | glmM | Core Essential | 3 |
| b0156 | erpA | Core Essential | 2 | b3729 | glmS | Core Essential | 5 |
| b0954 | fabA | Core Essential | 3 | b3730 | glmU | Core Essential | 5 |
| b2323 | fabB | Core Essential | 3 | b3870 | glnA | Conditional Essential | 1 |
| b1092 | fabD | Core Essential | 5 | b0680 | glnS | Core Essential | 5 |
| b1093 | fabG | Core Essential | 5 | b0720 | gltA | Core Essential | 3 |
| b1091 | fabH | Conditional Essential | 1 | b2400 | gltX | Core Essential | 5 |
| b1288 | fabI | Core Essential | 3 | b2551 | glyA | Core Essential | 3 |
| b0180 | fabZ | Core Essential | 3 | b3560 | glyQ | Core Essential | 4 |
| b2925 | fbaA | Core Essential | 3 | b3559 | glyS | Core Essential | 4 |
| b0592 | fepB | Conditional Essential | 1 | b3978 | glyT | Core Essential | 2 |
| b0588 | fepC | Conditional Essential | 1 | b3648 | gmk | Core Essential | 5 |

| | | | | | | | |
|-------|------|-----------------------|---|-------|------|-----------------------|---|
| b1550 | gnsB | Conditional Essential | 1 | b3772 | ilvA | Conditional Essential | 1 |
| b3608 | gpsA | Core Essential | 5 | b3774 | ilvC | Core Essential | 3 |
| b3181 | greA | Core Essential | 1 | b3771 | ilvD | Core Essential | 3 |
| b3406 | greB | Core Essential | 1 | b3770 | ilvE | Core Essential | 3 |
| b4143 | groL | Core Essential | 3 | b4669 | ilvX | Conditional Essential | 1 |
| b4142 | groS | Core Essential | 4 | N/A | imp | Core Essential | 1 |
| b2614 | grpE | Core Essential | 5 | b0884 | infA | Core Essential | 5 |
| b2507 | guaA | Core Essential | 3 | b3168 | infB | Core Essential | 5 |
| b2508 | guaB | Conditional Essential | 1 | b1718 | infC | Core Essential | 5 |
| b2231 | gyrA | Core Essential | 3 | b1160 | iraM | Conditional Essential | 1 |
| b3699 | gyrB | Core Essential | 3 | N/A | iroK | Conditional Essential | 1 |
| b2496 | hda | Core Essential | 1 | b2530 | iscS | Core Essential | 2 |
| b1210 | hemA | Core Essential | 5 | b0421 | ispA | Core Essential | 3 |
| b0369 | hemB | Core Essential | 5 | b3187 | ispB | Core Essential | 3 |
| b3805 | hemC | Core Essential | 5 | b2747 | ispD | Core Essential | 4 |
| b3804 | hemD | Core Essential | 4 | b1208 | ispE | Core Essential | 5 |
| b3997 | hemE | Conditional Essential | 1 | b2746 | ispF | Core Essential | 5 |
| b3850 | hemG | Core Essential | 3 | b2515 | ispG | Core Essential | 5 |
| b0475 | hemH | Core Essential | 3 | b0029 | ispH | Core Essential | 5 |
| N/A | hemK | Core Essential | 1 | b0174 | ispU | Core Essential | 3 |
| b0154 | hemL | Core Essential | 5 | b1215 | kdsA | Core Essential | 5 |
| b3082 | higA | Conditional Essential | 1 | b0918 | kdsB | Core Essential | 5 |
| b1508 | hipB | Conditional Essential | 1 | b3198 | kdsC | Core Essential | 1 |
| b2024 | hisA | Core Essential | 3 | N/A | kdtA | Core Essential | 1 |
| b2022 | hisB | Core Essential | 3 | b1352 | kilR | Conditional Essential | 1 |
| b2021 | hisC | Core Essential | 2 | b4421 | ldrB | Conditional Essential | 1 |
| b2020 | hisD | Core Essential | 3 | b2568 | lepB | Core Essential | 3 |
| b2025 | hisF | Core Essential | 3 | b0074 | leuA | Core Essential | 3 |
| b2019 | hisG | Core Essential | 3 | b0073 | leuB | Core Essential | 3 |
| b2023 | hisH | Core Essential | 3 | b0072 | leuC | Core Essential | 3 |
| b2026 | hisI | Core Essential | 3 | b0071 | leuD | Core Essential | 3 |
| b3797 | hisR | Core Essential | 2 | b0642 | leuS | Core Essential | 5 |
| b2514 | hisS | Core Essential | 5 | b3174 | leuU | Core Essential | 2 |
| N/A | hldD | Conditional Essential | 1 | b0672 | leuW | Core Essential | 2 |
| N/A | hldE | Conditional Essential | 1 | b1909 | leuZ | Core Essential | 2 |
| b1182 | hlyE | Conditional Essential | 1 | b4043 | lexA | Core Essential | 3 |
| b0640 | holA | Core Essential | 2 | b2828 | lgt | Core Essential | 3 |
| b1099 | holB | Core Essential | 2 | b2411 | ligA | Core Essential | 3 |
| b4259 | holC | Conditional Essential | 1 | b0628 | lipA | Conditional Essential | 1 |
| b4372 | holD | Conditional Essential | 1 | b0630 | lipB | Conditional Essential | 1 |
| b1842 | holE | Conditional Essential | 1 | b0657 | Int | Core Essential | 2 |
| b4350 | hsdR | Conditional Essential | 1 | b0891 | lolA | Core Essential | 3 |
| b1136 | icd | Core Essential | 2 | b1209 | lolB | Core Essential | 3 |
| b4519 | icdC | Conditional Essential | 1 | b1116 | lolC | Core Essential | 3 |
| b0026 | ileS | Core Essential | 4 | b1117 | lolD | Core Essential | 3 |

| | | | | | | | |
|-------|------|-----------------------|---|-------|------|-----------------------|---|
| b1118 | lolE | Core Essential | 3 | N/A | mrsA | Core Essential | 1 |
| b0116 | lpd | Conditional Essential | 1 | b0914 | msbA | Core Essential | 5 |
| b3200 | lptA | Core Essential | 4 | b0159 | mtn | Core Essential | 2 |
| b3201 | lptB | Core Essential | 3 | b0924 | mukB | Core Essential | 3 |
| b3199 | lptC | Core Essential | 3 | b0923 | mukE | Core Essential | 2 |
| b0054 | lptD | Core Essential | 2 | b0922 | mukF | Core Essential | 3 |
| b0641 | lptE | Core Essential | 2 | b3189 | murA | Core Essential | 5 |
| b4261 | lptF | Core Essential | 4 | b3972 | murB | Core Essential | 5 |
| b4262 | lptG | Core Essential | 4 | b0091 | murC | Core Essential | 5 |
| b0181 | lpxA | Core Essential | 5 | b0088 | murD | Core Essential | 5 |
| b0182 | lpxB | Core Essential | 5 | b0085 | murE | Core Essential | 5 |
| b0096 | lpxC | Core Essential | 4 | b0086 | murF | Core Essential | 5 |
| b0179 | lpxD | Core Essential | 5 | b0090 | murG | Core Essential | 5 |
| b0524 | lpxH | Core Essential | 5 | b3967 | murI | Core Essential | 4 |
| b0915 | lpxK | Core Essential | 5 | b1069 | murJ | Core Essential | 4 |
| b1054 | lpxL | Conditional Essential | 1 | N/A | mviN | Core Essential | 1 |
| b0027 | lspA | Core Essential | 3 | b0639 | nadD | Core Essential | 3 |
| b2687 | luxS | Core Essential | 1 | b1740 | nadE | Core Essential | 3 |
| b2838 | lysA | Core Essential | 3 | b2615 | nadK | Core Essential | 2 |
| b2890 | lysS | Conditional Essential | 1 | b2234 | nrdA | Core Essential | 3 |
| b0168 | map | Core Essential | 4 | b2235 | nrdB | Core Essential | 3 |
| N/A | mazE | Core Essential | 1 | b3169 | nusA | Core Essential | 5 |
| b4013 | metA | Core Essential | 3 | b0416 | nusB | Core Essential | 2 |
| b3939 | metB | Core Essential | 3 | b3982 | nusG | Core Essential | 5 |
| b3829 | metE | Conditional Essential | 1 | b3183 | obgE | Core Essential | 2 |
| b3941 | metF | Core Essential | 3 | b2082 | ogrK | Conditional Essential | 1 |
| b2114 | metG | Core Essential | 4 | b4162 | orn | Core Essential | 5 |
| b2942 | metK | Core Essential | 5 | b3019 | parC | Core Essential | 2 |
| b3828 | metR | Conditional Essential | 1 | b3030 | parE | Core Essential | 3 |
| b1114 | mfd | Core Essential | 1 | b0052 | pdxA | Core Essential | 2 |
| b4171 | miaA | Core Essential | 1 | b2320 | pdxB | Core Essential | 2 |
| b0661 | miaB | Core Essential | 1 | b1638 | pdxH | Conditional Essential | 1 |
| b1175 | minD | Core Essential | 1 | b2564 | pdxJ | Conditional Essential | 1 |
| b1174 | minE | Core Essential | 1 | b2926 | pgk | Core Essential | 3 |
| b1133 | mnmA | Core Essential | 1 | b1912 | pgsA | Core Essential | 3 |
| b2324 | mnmC | Core Essential | 1 | b2599 | pheA | Core Essential | 2 |
| b3706 | mnmE | Core Essential | 1 | b1715 | pheM | Conditional Essential | 1 |
| b3741 | mnmG | Core Essential | 1 | b1714 | pheS | Core Essential | 5 |
| b3021 | mqsA | Core Essential | 1 | b1713 | pheT | Core Essential | 5 |
| b0087 | mraY | Core Essential | 5 | b4041 | plsB | Core Essential | 3 |
| b0635 | mrda | Core Essential | 3 | b3018 | plsC | Core Essential | 4 |
| b0634 | mrdb | Core Essential | 3 | b3164 | pnp | Core Essential | 1 |
| b3251 | mreB | Core Essential | 3 | b3863 | polA | Core Essential | 2 |
| b3250 | mreC | Core Essential | 3 | b4226 | ppa | Core Essential | 3 |
| b3249 | mreD | Core Essential | 3 | b3956 | ppc | Conditional Essential | 1 |

| | | | | | | | |
|-------|------|-----------------------|---|-------|------|-----------------------|---|
| b1211 | prfA | Core Essential | 5 | b1277 | ribA | Core Essential | 5 |
| b2891 | prfB | Core Essential | 4 | b3041 | ribB | Core Essential | 4 |
| b3935 | priA | Core Essential | 2 | b1662 | ribC | Core Essential | 5 |
| b4201 | priB | Conditional Essential | 1 | b0414 | ribD | Core Essential | 5 |
| b1212 | prmC | Core Essential | 2 | b0415 | ribE | Core Essential | 3 |
| b0243 | proA | Conditional Essential | 1 | b0025 | ribF | Core Essential | 5 |
| b0242 | proB | Conditional Essential | 1 | b2608 | rimM | Core Essential | 2 |
| b0386 | proC | Core Essential | 3 | b3282 | rimN | Core Essential | 2 |
| b2189 | proL | Conditional Essential | 1 | b4180 | rlmB | Core Essential | 1 |
| b3799 | proM | Core Essential | 2 | b0859 | rlmC | Core Essential | 1 |
| b0194 | proS | Core Essential | 5 | b2785 | rlmD | Core Essential | 1 |
| b1207 | prs | Core Essential | 2 | b3179 | rlmE | Core Essential | 1 |
| N/A | prsA | Core Essential | 1 | b0807 | rlmF | Core Essential | 1 |
| b4160 | psd | Core Essential | 4 | b3084 | rlmG | Core Essential | 1 |
| b2585 | pssA | Core Essential | 5 | b0636 | rlmH | Core Essential | 1 |
| b1204 | pth | Core Essential | 2 | b0967 | rlmI | Core Essential | 1 |
| b2415 | ptsH | Conditional Essential | 1 | b0948 | rlmL | Core Essential | 1 |
| b2416 | ptsI | Conditional Essential | 1 | b2806 | rlmM | Core Essential | 1 |
| b4177 | purA | Core Essential | 3 | b2517 | rlmN | Core Essential | 1 |
| b1131 | purB | Core Essential | 4 | N/A | rlpB | Core Essential | 1 |
| b2476 | purC | Core Essential | 3 | b0058 | rluA | Core Essential | 1 |
| b4005 | purD | Core Essential | 3 | b1269 | rluB | Core Essential | 1 |
| b0523 | purE | Core Essential | 3 | b1086 | rluC | Core Essential | 1 |
| b2312 | purF | Core Essential | 3 | b2594 | rluD | Core Essential | 1 |
| b4006 | purH | Core Essential | 3 | b1135 | rluE | Core Essential | 1 |
| b0522 | purK | Core Essential | 3 | b4022 | rluF | Core Essential | 1 |
| b2557 | purL | Core Essential | 3 | b2567 | rnc | Core Essential | 2 |
| b2499 | purM | Core Essential | 3 | b1084 | rne | Core Essential | 4 |
| b4245 | pyrB | Core Essential | 2 | b3247 | rng | Core Essential | 1 |
| b1062 | pyrC | Core Essential | 3 | b3704 | rnpA | Core Essential | 5 |
| b3642 | pyrE | Core Essential | 3 | b3123 | rnpB | Core Essential | 2 |
| b1281 | pyrF | Core Essential | 3 | b1652 | rnt | Core Essential | 2 |
| b2780 | pyrG | Core Essential | 3 | b3386 | rpe | Conditional Essential | 1 |
| b0171 | pyrH | Core Essential | 3 | b2914 | rpiA | Conditional Essential | 1 |
| b0405 | queA | Core Essential | 1 | b3984 | rplA | Core Essential | 1 |
| b0444 | queC | Core Essential | 1 | b3317 | rplB | Core Essential | 5 |
| b2765 | queD | Core Essential | 1 | b3320 | rplC | Core Essential | 5 |
| b2794 | queF | Core Essential | 1 | b3319 | rplD | Core Essential | 5 |
| b1356 | racR | Core Essential | 2 | b3308 | rplE | Core Essential | 5 |
| b3167 | rbfA | Core Essential | 1 | b3305 | rplF | Core Essential | 5 |
| N/A | rcbA | Conditional Essential | 1 | b4203 | rplI | Core Essential | 1 |
| b3903 | rhaA | Conditional Essential | 1 | b3985 | rplJ | Core Essential | 5 |
| b3904 | rhaB | Conditional Essential | 1 | b3983 | rplK | Core Essential | 2 |
| b3780 | rhIB | Core Essential | 1 | b3986 | rplL | Core Essential | 5 |
| b3783 | rho | Core Essential | 4 | b3231 | rplM | Core Essential | 5 |

| | | | | | | | |
|-------|------|----------------|---|-------|------|-----------------------|---|
| b3310 | rpIN | Core Essential | 5 | b2609 | rpsP | Core Essential | 5 |
| b3301 | rpIO | Core Essential | 5 | b3311 | rpsQ | Core Essential | 4 |
| b3313 | rpIP | Core Essential | 5 | b4202 | rpsR | Core Essential | 4 |
| b3294 | rpIQ | Core Essential | 5 | b3316 | rpsS | Core Essential | 5 |
| b3304 | rpIR | Core Essential | 5 | b0023 | rpsT | Core Essential | 2 |
| b2606 | rpIS | Core Essential | 5 | b3065 | rpsU | Core Essential | 3 |
| b1716 | rpIT | Core Essential | 5 | b1822 | rrmA | Core Essential | 1 |
| b3186 | rpIU | Core Essential | 5 | b0176 | rseP | Core Essential | 2 |
| b3315 | rpIV | Core Essential | 5 | b4603 | rseX | Conditional Essential | 1 |
| b3318 | rpIW | Core Essential | 5 | b0051 | rsmA | Core Essential | 1 |
| b3309 | rpIX | Core Essential | 5 | b3289 | rsmB | Core Essential | 1 |
| b2185 | rpIY | Core Essential | 2 | b4371 | rsmC | Core Essential | 1 |
| b3185 | rpmA | Core Essential | 5 | b3465 | rsmD | Core Essential | 1 |
| b3637 | rpmB | Core Essential | 5 | b2946 | rsmE | Core Essential | 1 |
| b3312 | rpmC | Core Essential | 5 | b1835 | rsmF | Core Essential | 1 |
| b3302 | rpmD | Core Essential | 5 | b3740 | rsmG | Core Essential | 1 |
| b3936 | rpmE | Core Essential | 1 | b3146 | rsml | Core Essential | 1 |
| b1089 | rpmF | Core Essential | 1 | b2183 | rsuA | Core Essential | 1 |
| b3636 | rpmG | Core Essential | 1 | b4430 | rydB | Conditional Essential | 1 |
| b3703 | rpmH | Core Essential | 5 | b4597 | rydC | Conditional Essential | 1 |
| b1717 | rpml | Core Essential | 1 | N/A | safA | Conditional Essential | 1 |
| b3299 | rpmJ | Core Essential | 2 | b4082 | sdsR | Conditional Essential | 1 |
| b3295 | rpoA | Core Essential | 5 | b0098 | secA | Core Essential | 3 |
| b3987 | rpoB | Core Essential | 5 | b3609 | secB | Conditional Essential | 1 |
| b3988 | rpoC | Core Essential | 5 | b0408 | secD | Core Essential | 3 |
| b3067 | rpoD | Core Essential | 4 | b3981 | secE | Core Essential | 3 |
| b2573 | rpoE | Core Essential | 4 | b0409 | secF | Core Essential | 3 |
| b3461 | rpoH | Core Essential | 4 | b0097 | secM | Core Essential | 1 |
| b2741 | rpoS | Core Essential | 1 | b3300 | secY | Core Essential | 3 |
| b3649 | rpoZ | Core Essential | 1 | b2913 | serA | Conditional Essential | 1 |
| b0911 | rpsA | Core Essential | 5 | b4388 | serB | Conditional Essential | 1 |
| b0169 | rpsB | Core Essential | 5 | b0907 | serC | Core Essential | 3 |
| b3314 | rpsC | Core Essential | 5 | b0893 | serS | Core Essential | 5 |
| b3296 | rpsD | Core Essential | 5 | b0971 | serT | Core Essential | 2 |
| b3303 | rpsE | Core Essential | 5 | b2695 | serV | Core Essential | 2 |
| b4200 | rpsF | Core Essential | 2 | b3470 | sirA | Core Essential | 1 |
| b3341 | rpsG | Core Essential | 4 | b3650 | spoT | Core Essential | 2 |
| b3306 | rpsH | Core Essential | 5 | b1480 | sra | Core Essential | 1 |
| b3230 | rpsI | Core Essential | 4 | b4059 | ssb | Core Essential | 3 |
| b3321 | rpsJ | Core Essential | 5 | b0726 | sucA | Conditional Essential | 1 |
| b3297 | rpsK | Core Essential | 5 | b0727 | sucB | Conditional Essential | 1 |
| b3342 | rpsL | Core Essential | 4 | b2533 | suhB | Core Essential | 2 |
| b3298 | rpsM | Core Essential | 4 | b2559 | tadA | Core Essential | 4 |
| b3307 | rpsN | Core Essential | 5 | b3113 | tdcF | Core Essential | 1 |
| b3165 | rpsO | Core Essential | 1 | b4414 | tff | Conditional Essential | 1 |

| | | | | | | | |
|-------|------|-----------------------|---|-------|-------|-----------------------|---|
| b0406 | tgt | Core Essential | 1 | b4040 | ubiA | Core Essential | 3 |
| b3991 | thiG | Conditional Essential | 1 | b3835 | ubiB | Core Essential | 2 |
| b3990 | thiH | Core Essential | 1 | b3843 | ubiD | Core Essential | 2 |
| b0423 | thiI | Core Essential | 1 | b3833 | ubiE | Conditional Essential | 1 |
| b0417 | thiL | Core Essential | 5 | b2232 | ubiG | Conditional Essential | 1 |
| b0002 | thrA | Conditional Essential | 1 | b2311 | ubiX | Conditional Essential | 1 |
| b0003 | thrB | Core Essential | 3 | b4258 | valS | Core Essential | 5 |
| b0004 | thrC | Core Essential | 3 | b3633 | waaA | Core Essential | 3 |
| b1719 | thrS | Core Essential | 5 | N/A | waaU | Core Essential | 1 |
| b3976 | thrU | Core Essential | 2 | b3793 | wzyE | Core Essential | 2 |
| b2827 | thyA | Core Essential | 3 | N/A | yacA | Core Essential | 1 |
| b0436 | tig | Core Essential | 1 | N/A | yacE | Core Essential | 1 |
| b0188 | tilS | Core Essential | 5 | N/A | yadF | Core Essential | 1 |
| b2474 | tmcA | Core Essential | 1 | N/A | yadR | Core Essential | 1 |
| b1098 | tmk | Core Essential | 5 | N/A | yaeL | Core Essential | 1 |
| b3709 | tnaB | Core Essential | 1 | N/A | yaeT | Core Essential | 1 |
| b1252 | tonB | Conditional Essential | 1 | b0232 | yafN | Core Essential | 1 |
| b1274 | topA | Core Essential | 2 | b0270 | yagG | Core Essential | 1 |
| b3919 | tpiA | Conditional Essential | 1 | b0498 | ybbC | Conditional Essential | 1 |
| b3965 | trmA | Core Essential | 1 | b0704 | ybfC | Conditional Essential | 1 |
| b2607 | trmD | Core Essential | 5 | b0969 | ycckK | Core Essential | 1 |
| b3651 | trmH | Core Essential | 1 | b1085 | yceQ | Core Essential | 2 |
| b2960 | trmI | Core Essential | 1 | N/A | ycfB | Core Essential | 1 |
| b2532 | trmJ | Core Essential | 1 | b1254 | yciB | Conditional Essential | 1 |
| b1260 | trpA | Core Essential | 2 | b1259 | yciG | Conditional Essential | 1 |
| b1261 | trpB | Core Essential | 3 | b1280 | yciM | Conditional Essential | 1 |
| b1262 | trpC | Core Essential | 2 | b4526 | ydaE | Conditional Essential | 1 |
| b1263 | trpD | Core Essential | 2 | b1457 | ydcD | Conditional Essential | 1 |
| b1264 | trpE | Core Essential | 2 | b1472 | yddL | Conditional Essential | 1 |
| b1265 | trpL | Conditional Essential | 1 | b1499 | ydeO | Conditional Essential | 1 |
| b3384 | trpS | Core Essential | 5 | b1505 | ydeT | Conditional Essential | 1 |
| b3761 | trpT | Core Essential | 1 | b1572 | ydfB | Core Essential | 1 |
| b2318 | truA | Core Essential | 1 | b1705 | ydiE | Conditional Essential | 1 |
| b3166 | truB | Core Essential | 1 | b1689 | ydiL | Core Essential | 1 |
| b2791 | truC | Core Essential | 1 | b1807 | yeaZ | Core Essential | 1 |
| b2745 | truD | Core Essential | 1 | N/A | yedN | Conditional Essential | 1 |
| N/A | tsaB | Core Essential | 1 | b2017 | yefM | Core Essential | 1 |
| N/A | tsaC | Core Essential | 2 | b2110 | yehC | Conditional Essential | 1 |
| N/A | tsaD | Core Essential | 2 | b2111 | yehD | Conditional Essential | 1 |
| N/A | tsaE | Core Essential | 2 | b2188 | yejM | Core Essential | 3 |
| b0170 | tsf | Core Essential | 5 | b2238 | yfaH | Conditional Essential | 1 |
| b1344 | ttcA | Core Essential | 1 | b2273 | yfbN | Conditional Essential | 1 |
| N/A | ttcC | Conditional Essential | 1 | b2450 | yffS | Conditional Essential | 1 |
| b2600 | tyrA | Core Essential | 3 | b2575 | yfiC | Core Essential | 1 |
| b1637 | tyrS | Core Essential | 5 | N/A | yfiO | Core Essential | 1 |

| | | | | | | | |
|-------|------|-----------------------|---|-------|------|-----------------------|---|
| N/A | yfjA | Core Essential | 1 | N/A | yjgP | Core Essential | 1 |
| N/A | yfjB | Core Essential | 1 | N/A | yjgQ | Core Essential | 1 |
| b2850 | ygeF | Conditional Essential | 1 | b4279 | yjhB | Conditional Essential | 1 |
| b2851 | ygeG | Conditional Essential | 1 | b4586 | ykfM | Conditional Essential | 1 |
| b2855 | ygeK | Conditional Essential | 1 | b4671 | ykgR | Conditional Essential | 1 |
| b2858 | ygeN | Conditional Essential | 1 | b1137 | ymfD | Conditional Essential | 1 |
| b2898 | ygfZ | Conditional Essential | 1 | b1138 | ymfE | Conditional Essential | 1 |
| N/A | ygiG | Core Essential | 1 | b1167 | ymgC | Conditional Essential | 1 |
| b3064 | yjjD | Core Essential | 1 | N/A | ymiB | Conditional Essential | 1 |
| N/A | yhbG | Core Essential | 1 | b4673 | ymjD | Conditional Essential | 1 |
| N/A | yhbN | Core Essential | 1 | b4674 | ynbG | Conditional Essential | 1 |
| b3159 | yhbV | Core Essential | 1 | b1436 | yncJ | Conditional Essential | 1 |
| b3343 | yheL | Core Essential | 1 | b1788 | yoal | Conditional Essential | 1 |
| b3344 | yheM | Core Essential | 1 | b4677 | yobl | Conditional Essential | 1 |
| b3345 | yheN | Core Essential | 1 | b4543 | ypaA | Conditional Essential | 1 |
| b3471 | yhhQ | Core Essential | 1 | b4682 | yqcG | Conditional Essential | 1 |
| b3705 | yidC | Core Essential | 3 | b2848 | yqeJ | Conditional Essential | 1 |
| b3865 | yihA | Core Essential | 2 | b2949 | yqgF | Core Essential | 2 |
| b4047 | yjbL | Conditional Essential | 1 | N/A | yrbl | Core Essential | 1 |
| b4048 | yjbM | Conditional Essential | 1 | N/A | yrbK | Core Essential | 1 |
| b4621 | yjbS | Conditional Essential | 1 | N/A | yrdC | Core Essential | 1 |
| b4168 | yjeE | Core Essential | 1 | b3398 | yrfF | Core Essential | 3 |
| N/A | yjeQ | Core Essential | 1 | b2412 | zipA | Core Essential | 3 |
| b4166 | yjeS | Core Essential | 1 | | | | |

Tabla 2 | Información de los TFs candidatos definidos por ReProMin. Número de objetivos silenciados e inducidos no esenciales para los 34 TFs candidatos definidos.

| TF | Silenced Genes | Induced Genes |
|---------------|-----------------------|----------------------|
| CueR | 2 | 0 |
| PhoB | 27 | 4 |
| FlhDC | 18 | 0 |
| MarA | 2 | 0 |
| CsgD | 6 | 3 |
| AsnC | 1 | 0 |
| YqhC | 2 | 0 |
| CreB | 1 | 0 |
| DhaR | 3 | 0 |
| RbsR | 1 | 0 |
| YdeO | 2 | 0 |
| SgrR | 1 | 4 |
| Nac | 5 | 1 |
| RcdA | 2 | 1 |
| ZntR | 1 | 0 |
| RclR | 3 | 0 |
| MalT | 1 | 0 |
| ZraR | 1 | 0 |
| AlIS | 3 | 0 |
| RcsAB | 5 | 0 |
| Ada | 1 | 0 |
| CynR | 3 | 0 |
| Dan | 4 | 0 |
| FeaR | 1 | 0 |
| GcvA | 1 | 0 |
| HcaR | 5 | 0 |
| KdpE | 4 | 0 |
| MatA | 2 | 0 |
| RcsB- BglJ | 11 | 0 |
| Rob | 1 | 0 |
| XapR | 2 | 0 |
| YpdB | 1 | 0 |
| OmpR | 3 | 1 |
| IscR | 1 | 6 |

Tabla 3 | Resumen del análisis combinatorio ReProMin de los casos de ST y UT. #KO: Tamaño de la combinación; #TFs: Número de TFs a incluir en el análisis; #Combinations: número total de combinaciones a probar; Time: cálculo del tiempo empleado en resolver todas las combinaciones; comb/seg: combinaciones resueltas por segundo.

Caso ST

| #KO | #TFs | #Combinations | Time (hrs) | comb/seg |
|-----|------|---------------|------------|----------|
| 1 | 132 | 132 | 0 | 26890.48 |
| 2 | 132 | 8646 | 0 | 67223.19 |
| 3 | 132 | 374660 | 0 | 49280.48 |
| 4 | 132 | 12082785 | 0.09 | 36109.94 |
| 5 | 132 | 309319296 | 3.11 | 27644.81 |
| 6 | 132 | 6547258432 | 94.45 | 19254.84 |
| 7 | 132 | 1.17851E+11 | #N/A | #N/A |
| 127 | 132 | 309319296 | #N/A | #N/A |
| 128 | 132 | 12082785 | 40.55 | 82.78 |
| 129 | 132 | 374660 | 1.23 | 84.88 |
| 130 | 132 | 8646 | 0.04 | 66.55 |
| 131 | 132 | 132 | 0 | 28.37 |
| 132 | 132 | 1 | 0 | 83.76 |

Caso UT

| #KO | #TFs | #Combinations | Time (hrs) | comb/seg |
|-----|------|---------------|------------|----------|
| 1 | 20 | 20 | 0 | 31895.85 |
| 2 | 20 | 190 | 0 | 16578.28 |
| 3 | 20 | 1140 | 0 | 13641.57 |
| 4 | 20 | 4845 | 0 | 7722.92 |
| 5 | 20 | 15504 | 0 | 5664.81 |
| 6 | 20 | 38760 | 0 | 4464.86 |
| 7 | 20 | 77520 | 0.01 | 3328.11 |
| 8 | 20 | 125970 | 0.01 | 2811.83 |
| 9 | 20 | 167960 | 0.02 | 2402.82 |
| 10 | 20 | 184756 | 0.03 | 1933.12 |
| 11 | 20 | 167960 | 0.03 | 1759.93 |
| 12 | 20 | 125970 | 0.02 | 1444.05 |
| 13 | 20 | 77520 | 0.02 | 1278.03 |
| 14 | 20 | 38760 | 0.01 | 1197.04 |
| 15 | 20 | 15504 | 0 | 1028.29 |
| 16 | 20 | 4845 | 0 | 859.83 |
| 17 | 20 | 1140 | 0 | 928.78 |
| 18 | 20 | 190 | 0 | 698.71 |
| 19 | 20 | 20 | 0 | 721.36 |
| 20 | 20 | 1 | 0 | 601.16 |

Tabla 4 | Datos RNA-seq y proteómicos estimados para los objetivos ReProMin de PFC y los off-targets.

| ID | Name | Regulator | Regulation | WT_FPKM | PFC_FPKM | log2(fold_change) | p_value | Prediction_Accuracy | MW | T_Rate | WT_est_copies |
|-------|------|-----------|------------|-----------|-----------|-------------------|----------|---------------------|-------------|-------------|---------------|
| b0123 | cueO | cueR | + | 26.1335 | 3.59261 | -2.86279 | 5.00E-05 | 1 | 56556 | 39.5660742 | 1034 |
| b0484 | copA | cueR | + | 44.0292 | 16.1397 | -1.44784 | 5.00E-05 | 1 | 87799.79865 | 62.89008204 | 2769 |
| b1073 | flgB | flhC | + | 1.78268 | 0 | -1000 | 0.00075 | 1 | 15240 | 131.9226153 | 235.1758079 |
| b1074 | flgC | flhC | + | 1.39679 | 0 | -1000 | 0.00365 | 1 | 13968 | 131.9226153 | 184.2681899 |
| b1076 | flgE | flhC | + | 0.226069 | 0 | -1000 | 1 | 1 | 42045 | 176.9371298 | 40 |
| b1078 | flgG | flhC | + | 0.188845 | 0 | -1000 | 1 | 1 | 27744 | 131.9226153 | 24.91292629 |
| b1077 | flgF | flhC | + | 0.295374 | 0.131503 | -1.16745 | 1 | 1 | 25912 | 131.9226153 | 38.96651058 |
| b1079 | flgH | flhC | + | 0.216761 | 0.140042 | -0.630247 | 1 | 1 | 24615 | 131.9226153 | 28.59567802 |
| b1081 | flgJ | flhC | + | 0.150049 | 0.0972888 | -0.625085 | 1 | 1 | 34475 | 131.9226153 | 19.79485651 |
| b1075 | flgD | flhC | + | 0.438495 | 0.286657 | -0.613236 | 1 | 1 | 23575 | 131.9226153 | 57.84740721 |
| b1080 | flgI | flhC | + | 0.504011 | 0.331207 | -0.605721 | 1 | 1 | 38169 | 131.9226153 | 66.49044928 |
| b1072 | flgA | flhC | + | 1.63939 | 1.22695 | -0.41809 | 0.5387 | 1 | 23519 | 131.9226153 | 216.2726163 |
| b0655 | gltI | flhC | + | 4.96143 | 3.82443 | -0.375513 | 0.3366 | 1 | 33420 | 994.8744616 | 4936 |
| b1880 | flhB | flhC | + | 0.299985 | 0.237155 | -0.339058 | 1 | 1 | 42238 | 131.9226153 | 39.57480576 |
| b0652 | gltL | flhC | + | 13.794 | 13.7489 | -0.00473277 | 0.9834 | 1 | 26661 | 5.9446136 | 82 |
| b0654 | gltJ | flhC | + | 1.21607 | 1.32954 | 0.128695 | 1 | 0 | 27503 | 131.9226153 | 160.4271348 |
| b0653 | gltK | flhC | + | 0.682604 | 0.752963 | 0.14153 | 1 | 0 | 24915 | 131.9226153 | 90.05090492 |
| b1879 | flhA | flhC | + | 0.747208 | 0.859931 | 0.202712 | 1 | 0 | 74843 | 131.9226153 | 98.57363356 |
| b4763 | sroC | flhC | + | 0 | 0 | 0 | 1 | #N/A | 54678 | 131.9226153 | 0 |
| b4099 | phnI | phoB | + | 0.0658692 | 0 | -1000 | 1 | 1 | 38853 | 131.9226153 | 8.689637134 |
| b4096 | phnL | phoB | + | 0.112072 | 0 | -1000 | 1 | 1 | 24705 | 131.9226153 | 14.78483135 |
| b0400 | phoR | phoB | + | 1.72593 | 0.483428 | -1.836 | 0.0198 | 1 | 49629 | 131.9226153 | 227.6891995 |
| b4094 | phnN | phoB | + | 0.439148 | 0.194714 | -1.17335 | 1 | 1 | 20730 | 131.9226153 | 57.93355268 |
| b0384 | psiF | phoB | + | 6.57898 | 3.61509 | -0.863832 | 0.2013 | 1 | 11687 | 6.231969089 | 41 |
| b1982 | amn | phoB | + | 3.30993 | 2.16787 | -0.610524 | 0.11925 | 1 | 53995 | 128.7036282 | 426 |
| b4105 | phnD | phoB | + | 0.138975 | 0.0920939 | -0.59365 | 1 | 1 | 37371 | 131.9226153 | 18.33394547 |

| | | | | | | | | | | | |
|-------|-------|------|------|-----------|----------|------------|----------|------|--------|-------------|-------------|
| b4092 | phnP | phoB | + | 2.54763 | 1.80488 | -0.49725 | 0.3687 | 1 | 27848 | 20.01860553 | 51 |
| b2465 | tktB | phoB | + | 10.1271 | 8.10833 | -0.320744 | 0.2383 | 1 | 73043 | 74.45369355 | 754 |
| b1782 | mipA | phoB | + | 7.59519 | 6.53516 | -0.216865 | 0.55925 | 1 | 27831 | 438.1720536 | 3328 |
| b2987 | pitB | phoB | - | 0.266126 | 0.234914 | -0.179976 | 1 | 0 | 53809 | 131.9226153 | 35.10803793 |
| b0397 | sbcC | phoB | - | 2.29386 | 2.07824 | -0.142414 | 0.6811 | 0 | 118721 | 3.051624772 | 7 |
| b4106 | phnC | phoB | + | 0 | 0 | 0 | 1 | #N/A | 29431 | 131.9226153 | 0 |
| b4102 | phnF | phoB | + | 0 | 0 | 0 | 1 | #N/A | 27621 | 131.9226153 | 0 |
| b4101 | phnG | phoB | + | 0 | 0 | 0 | 1 | #N/A | 16540 | 131.9226153 | 0 |
| b4100 | phnH | phoB | + | 0 | 0 | 0 | 1 | #N/A | 21027 | 131.9226153 | 0 |
| b4098 | phnJ | phoB | + | 0 | 0 | 0 | 1 | #N/A | 31845 | 131.9226153 | 0 |
| b4095 | phnM | phoB | + | 0 | 0 | 0 | 1 | #N/A | 42010 | 131.9226153 | 0 |
| b4583 | phnE | phoB | + | 0 | 0 | 0 | 1 | #N/A | 0 | 131.9226153 | 0 |
| b1540 | ydfH | phoB | + | 4.78016 | 4.36114 | -0.132356 | 0.75665 | 1 | 26565 | 20.08300977 | 96 |
| b0080 | cra | phoB | - | 9.8679 | 9.4407 | -0.063849 | 0.8382 | 0 | 37999 | 14.99812523 | 148 |
| b0383 | phoA | phoB | + | 1.515 | 1.48745 | -0.0264802 | 0.95585 | 1 | 49439 | 131.9226153 | 199.8627622 |
| b4093 | phnO | phoB | + | 3.96923 | 4.08709 | 0.0422132 | 0.94135 | 0 | 16570 | 85.65893133 | 340 |
| b0398 | sbcD | phoB | - | 0.682064 | 0.737894 | 0.113508 | 1 | 1 | 44714 | 16.12751883 | 11 |
| b3521 | yhjC | phoB | + | 1.35302 | 1.46734 | 0.117019 | 0.82765 | 0 | 33330 | 131.9226153 | 178.493937 |
| b0241 | phoE | phoB | + | 0.795344 | 0.954383 | 0.26299 | 1 | 0 | 38922 | 131.9226153 | 104.9238606 |
| b1020 | phoH | phoB | + | 1.04242 | 1.29115 | 0.308721 | 1 | 0 | 39265 | 131.9226153 | 137.5187727 |
| b4097 | phnK | phoB | + | 0.0976597 | 0.261719 | 1.42218 | 1 | 0 | 27831 | 131.9226153 | 12.88352304 |
| b0487 | cueR | TF | #N/A | 7.37987 | 0 | -1000 | 5.00E-05 | #N/A | 15235 | 6.639683355 | 49 |
| b1891 | filhC | TF | #N/A | 0.555941 | 0 | -1000 | 1 | #N/A | 21566 | 131.9226153 | 73.34119069 |
| b0399 | phoB | TF | #N/A | 1.77246 | 0 | -1000 | 5.00E-05 | #N/A | 26433 | 3.949313384 | 7 |
| b3501 | arsR | #N/A | #N/A | 1.46292 | 0 | -1000 | 0.00955 | #N/A | 13253 | 131.9226153 | 192.9922324 |
| b1085 | yceQ | #N/A | #N/A | 2.08564 | 0 | -1000 | 0.003 | #N/A | 12100 | 131.9226153 | 275.1430834 |
| b2518 | ndk | #N/A | #N/A | 11.2188 | 2.74682 | -2.03008 | 0.00475 | #N/A | 15463 | 1039.772525 | 11665 |
| b4151 | frdD | #N/A | #N/A | 20.806 | 8.45121 | -1.29977 | 0.00445 | #N/A | 13107 | 131.9226153 | 2744.781935 |
| b3724 | phoU | #N/A | #N/A | 14.7371 | 6.11853 | -1.2682 | 0.00085 | #N/A | 27417 | 22.93531292 | 338 |
| b3261 | fis | #N/A | #N/A | 33.955 | 15.195 | -1.16003 | 0.00685 | #N/A | 11240 | 103.7549698 | 3523 |

| | | | | | | | | | | | |
|-------|------|------|------|---------|---------|----------|----------|------|--------|-------------|-------|
| b0884 | infA | #N/A | #N/A | 69.2935 | 31.0307 | -1.15902 | 0.00615 | #N/A | 8250 | 64.86899926 | 4495 |
| b2499 | purM | #N/A | #N/A | 12.9641 | 5.82321 | -1.15463 | 0.00145 | #N/A | 36854 | 196.234216 | 2544 |
| b0574 | cusB | #N/A | #N/A | 4.98823 | 9.99873 | 1.00322 | 0.0055 | #N/A | 44305 | 467.7009681 | 2333 |
| b1846 | yebE | #N/A | #N/A | 5.15941 | 10.7358 | 1.05715 | 0.0099 | #N/A | 23687 | 93.03389341 | 480 |
| b0572 | cusC | #N/A | #N/A | 268.268 | 563.755 | 1.07139 | 0.00075 | #N/A | 50270 | 5.49823311 | 1475 |
| b0575 | cusA | #N/A | #N/A | 5.18742 | 11.9441 | 1.20321 | 5.00E-05 | #N/A | 114707 | 68.62756438 | 356 |
| b0573 | cusF | #N/A | #N/A | 12.2061 | 34.2463 | 1.48835 | 0.0012 | #N/A | 12251 | 2418.217121 | 29517 |
| b4484 | cpxP | #N/A | #N/A | 4.75536 | 13.605 | 1.51651 | 0.0019 | #N/A | 18965 | 68.97479896 | 328 |
| b0466 | ybaM | #N/A | #N/A | 0 | 18.7291 | 1000 | 0.00025 | #N/A | 6025 | 131.9226153 | 0 |
| b0803 | ybil | #N/A | #N/A | 0 | 2.60701 | 1000 | 0.00695 | #N/A | 9756 | 131.9226153 | 0 |
| b1161 | ycgX | #N/A | #N/A | 0 | 1.81556 | 1000 | 0.0006 | #N/A | 15769 | 131.9226153 | 0 |

Tabla 5 | Cepas, Primers y Plásmidos utilizados en el estudio.

CEPAS

| Name | Characteristics | Reference |
|---------|---|---------------------------|
| BW25113 | F ⁺ , DE(araD-araB)567, lacZ4787(del)::rrnB-3, λ, rph-1, DE(rhaD-rhaB)568, hsdR514 | Baba <i>et al.</i> (2006) |
| FOG | BW25113 derivative. ΔfliA - ΔoxyR - ΔgadE | This study |
| PYC | BW25113 derivative. ΔphoB - ΔyedW - ΔcusR | This study |
| PFC | BW25113 derivative. ΔphoB - ΔflhC - ΔcueR | This study |

PRIMERS

| Name | Sequence 5' -> 3' | Purpose | Reference |
|--------|------------------------|--|------------|
| flhC_F | AAGTTGGTTATTCTGGGTGGGA | Forward primer for verification of flhC deletion | This study |
| flhC_R | CAAACCGCACCAATGTCCAG | Reverse primer for verification of flhC deletion | This study |
| phoB_F | CCACCAGATGACAGCCAGTT | Forward primer for verification of phoB deletion | This study |
| phoB_R | GCCGTATAGTAGCGGTTCCC | Reverse primer for verification of phoB deletion | This study |
| yedW_F | GTTTCTTGAAATAGTCCCCGGT | Forward primer for verification of yedW deletion | This study |
| yedW_R | CAGCTGCGTGATTGTCTCA | Reverse primer for verification of yedW deletion | This study |
| cusR_F | TCAGACGATATTGCGCCCG | Forward primer for verification of cusR deletion | This study |
| cusR_R | TATGGCGTCACGCGTAAACT | Reverse primer for verification of cusR deletion | This study |
| cueR_F | TTGATTTGCTACCTGGCGGT | Forward primer for verification of cueR deletion | This study |
| cueR_R | ATGCTGGGCGGAAATGGTTA | Reverse primer for verification of cueR deletion | This study |
| fliA_F | CCAAATTGCCTGATGCGCTG | Forward primer for verification of fliA deletion | This study |
| fliA_R | ACGACATATTCACGCACCGT | Reverse primer for verification of fliA deletion | This study |
| evgA_F | CAGGAACGTTGGAGGCCTAA | Forward primer for verification of evgA deletion | This study |
| evgA_R | AATACCACGAACCCGTTGCT | Reverse primer for verification of evgA deletion | This study |
| oxyR_F | ATCCGATTCTGTCGCTGCAA | Forward primer for verification of oxyR deletion | This study |
| oxyR_R | CTGTGGGCAAAACCGAACAG | Reverse primer for verification of oxyR deletion | This study |

| | | | |
|----|----------------------|---|--------------------------|
| K1 | CAGTCATAGCCGAATAGCCT | For verification of insertion of KanR. Binds within Kan cassette. Reverse primer. | Datsenko & Wanner (2000) |
|----|----------------------|---|--------------------------|

PLASMIDOS

| Name | Description | Reference |
|--------------|---|----------------------------------|
| pSEVA631 | Gmr , ori-pBBR1 | Silva-Rocha <i>et al.</i> (2013) |
| pSEVA63-Dual | pSEVA631 carrying the circuit MBP 1.0 | Darlington <i>et al.</i> (2013) |
| pSEVA63-Hvio | Vector carrying a set of genes enabling production of violacein. The circuit is composed with the vioA gene, constitutively expressed, and the vioBCDE genes, which are inducible | Darlington <i>et al.</i> (2013) |
| pE-FLP | Flp recombinase expression plasmid. Used to remove Km cassette resistance. Addgene #45978 | St-Pierre <i>et al.</i> (2013) |

- 1 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).
- 2 Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, 695 (2018).
- 3 St-Pierre, F. *et al.* One-Step Cloning and Chromosomal Integration of DNA. *ACS Synth. Biol.* **2**, 537–541 (2013).
- 4 Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, D666–D675 (2013).



A quantitative method for proteome reallocation using minimal regulatory interventions

Gustavo Lastiri-Pancardo¹, Jonathan S. Mercado-Hernández¹, Juhyun Kim², José I. Jiménez² and José Utrilla¹✉

Engineering resource allocation in biological systems is an ongoing challenge. Organisms allocate resources for ensuring survival, reducing the productivity of synthetic biology functions. Here we present a new approach for engineering the resource allocation of *Escherichia coli* by rationally modifying its transcriptional regulatory network. Our method (ReProMin) identifies the minimal set of genetic interventions that maximizes the savings in cell resources. To this end, we categorized transcription factors according to the essentiality of its targets and we used proteomic data to rank them. We designed the combinatorial removal of transcription factors that maximize the release of resources. Our resulting strain containing only three mutations, theoretically releasing 0.5% of its proteome, had higher proteome budget, increased production of an engineered metabolic pathway and showed that the regulatory interventions are highly specific. This approach shows that combining proteomic and regulatory data is an effective way of optimizing strains using conventional molecular methods.

The removal of accessory nonessential functions is one of the strategies used to engineer microbial phenotypes. This approach relies on the assumption that cellular resources for gene expression are limited and, therefore, by removing genes unneeded in a certain environment, the cell is capable of allocating resources to other functions (for example, expression of recombinant genes). These minimization approaches are mostly done by reducing genome size and gene number including performing random deletions^{1,2}, however, the precise way in which the resource allocation takes place after the genetic intervention is not considered.

Organisms respond to the environment by cellular signaling pathways encoded in regulatory networks³. The intricacy of the lifestyle of an organism is generally translated into signaling complexity⁴. Biological regulatory networks are robust⁵ and evolvable⁶ to cope with environmental and lifestyle perturbations, however, this robustness involves intrinsic trade-offs, such as resource allocation strategies. It has been shown that cellular states are naturally 'primed' for typical upcoming changes. Bacteria anticipate to fluctuations in the environment^{7,8} draining resources from functions that are mostly performed in relatively stable conditions. The expression of anticipation functions, also called hedging functions, is encoded in the regulatory network and has a proteomic cost⁹. Genome-scale models along with experimental datasets enable the calculation of the minimal theoretical proteome needed to sustain growth in a defined condition¹⁰. Therefore, comparing minimal theoretical proteomes with measured proteomes reveals the costs of the hedging proteome allocation. Proteome econometric approaches can facilitate the engineering of cellular states or phenotypes aimed at displaying an engineered function. Recent studies have focused on the host-construct interactions for increasing predictability of synthetic constructs¹¹⁻¹³. In addition to these approaches, the rational design of the host used for expression following econometric models can be adopted to improve the performance of synthetic constructs, including production phenotypes for molecules of added value. Among other benefits, streamlined organisms obtained this way are less likely to develop undesired emerging behaviors¹⁴.

We foresee the use of regulatory mechanism as a control layer that will aid in the design of cellular phenotypes. Our ability to engineer biological systems depends on understanding how cells sense and respond to their environment at a system level. Few studies have tackled this issue and none of them in a rational way.

In this work we developed a new top-down cell engineering strategy for *E. coli* using the transcriptional regulatory network (TRN) as a control layer for proteome allocation. By combining high-throughput proteomic information, regulatory network interactions and gene essentiality observations, we implemented a method capable of finding the minimal set of genetic interventions required to divert the resources invested in superfluous hedging into increased biosynthetic potential. The resulting strains exhibited an increased availability of cellular resources to express engineered functions.

Results

Identification of dispensable TFs for proteome release. The genome-scale model of metabolism and gene expression (ME-model) computes the minimal theoretical proteome and allows calculating the cost of expressing hedging functions. It can be used to simulate different scenarios of expression of the hedging proteome (as unused protein fraction coefficient, see Methods)¹⁴. These simulations allowed us to calculate the costs and the benefits of different interventions, for example by modulating the expression of the hedging proteome, expressed in terms of growth, the size of both essential and recombinant proteome sectors (Extended Data Fig. 1).

We built on ME-models to design strains containing the minimal genetic interventions that reduced the greatest amount of proteomic resources not required to grow in a specific condition. Our method used transcription factors (TFs) as the key dials controlling the allocation of the hedging proteome in a predefined specific environment. We began by establishing batch growth in minimal medium (M9) supplemented with glucose as the sole carbon source as the defined environment for the first case of this study. Then, by compiling experimental and genome-scale model generated essential gene sets, we generated a list of essential genes for growth in this specific

¹Systems and Synthetic Biology Program, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico.

²Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK. ✉e-mail: utrilla@ccg.unam.mx

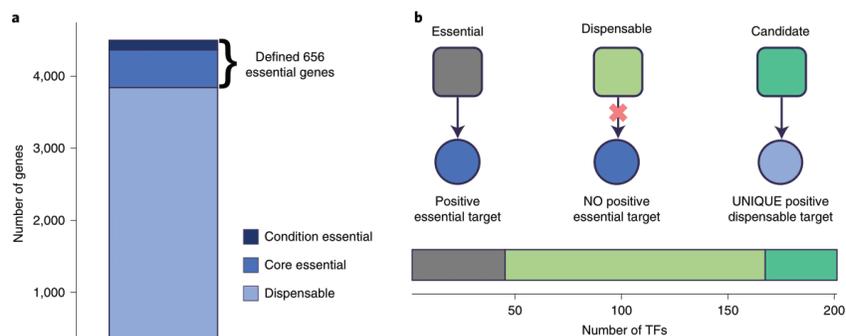


Fig. 1 | Gene essentiality and TRN analysis in the predefined condition. **a**, Essentiality profile of the *E. coli* genes (~4,600) under selected growth condition. The essential genes considered for this analysis are divided into core essential (523 genes, always needed) and conditional (133 genes, M9-glucose needed). **b**, Graphical representation of the subnetwork of TF-gene interactions considered for the classification of the TFs; gray squares represent essential TFs, light green squares dispensable TFs, dark green squares candidate TFs, dark blue circles essential genes, light blue circles dispensable genes and arrows positive interactions. The lower bar shows the number of TFs in each classification from the total considered (203 TFs).

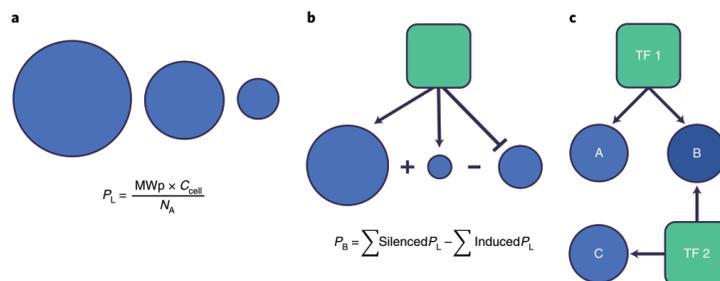


Fig. 2 | Emerging properties from proteomics data integration. **a**, The P_L of a gene is defined as the molecular weight of the protein (MW_p) multiplied by the number of copies per cell (C_{cell}) divided by Avogadro's number (N_A) (6.022×10^{23} fg equivalent). The more expressed the gene is, the more P_L it generates. **b**, The P_B of a TF defined as the sum of the P_L of the silenced genes minus the sum of the P_L of the induced genes. **c**, Schematic of a simple case of shared regulation in which removing both TFs silences all target genes but this is not the case when the TFs are silenced individually.

environment (Fig. 1a and Supplementary Table 1, see Methods). Once the case-specific gene essentiality was defined, we analyzed the TF-gene interactions compiled in RegulonDB¹⁵. After determining gene essentiality and TF-gene regulatory interactions, we analyzed the subnetwork of interactions of each TF looking for dispensable TFs, defined as those that do not activate the expression of any essential gene. According to our analysis, 156 of the 200 TFs contained in the regulatory network can be eliminated (Fig. 1b). Since our goal was to reduce the hedging proteome, out of the 156 dispensable TFs we selected as candidates for nonessential function reduction the 34 TFs with at least one unique (meaning it is not activated by any other TF) positive regulated gene (Supplementary Table 2) (see Methods); this gives the certainty of silencing at least one gene.

Integration of proteomic and regulatory network data. We determined the proteome associated to each nonessential TF in our network integrating a quantitative proteomic dataset¹⁶ that provides protein copy number per cell under 22 different growth conditions with 95% of proteome coverage (by mass). Here we defined two

emerging properties derived from the quantitative proteomics data integration: the proteomic load of a gene (P_L) in femtograms (fg) of protein per cell (Fig. 2a) and the proteomic balance (P_B) of a TF resulting from the summation of the P_L of the genes that would result silenced or activated by the elimination of a TF (Fig. 2b). P_B is conceptually important to rank the TFs according to the size of the proteome they control, since it considers the net addition of protein mass (in fg of protein per cell) liberated when removing a TF.

Computational search of minimal TF eliminations. Many TFs have shared target genes (Fig. 2c), in fact, many of them are part of a simplified version of a dense overlapping regulon networks motif¹⁷, meaning that a particular combination of TFs is needed to 'fully' silence these targets, generating a full landscape of potential proteome liberation composed by all the different combinatorial TFs deletions.

To assist with the design of the mutant strains we developed a computational method. We called our method ReProMin (regulation-based proteome minimization); it integrates the TF-gene

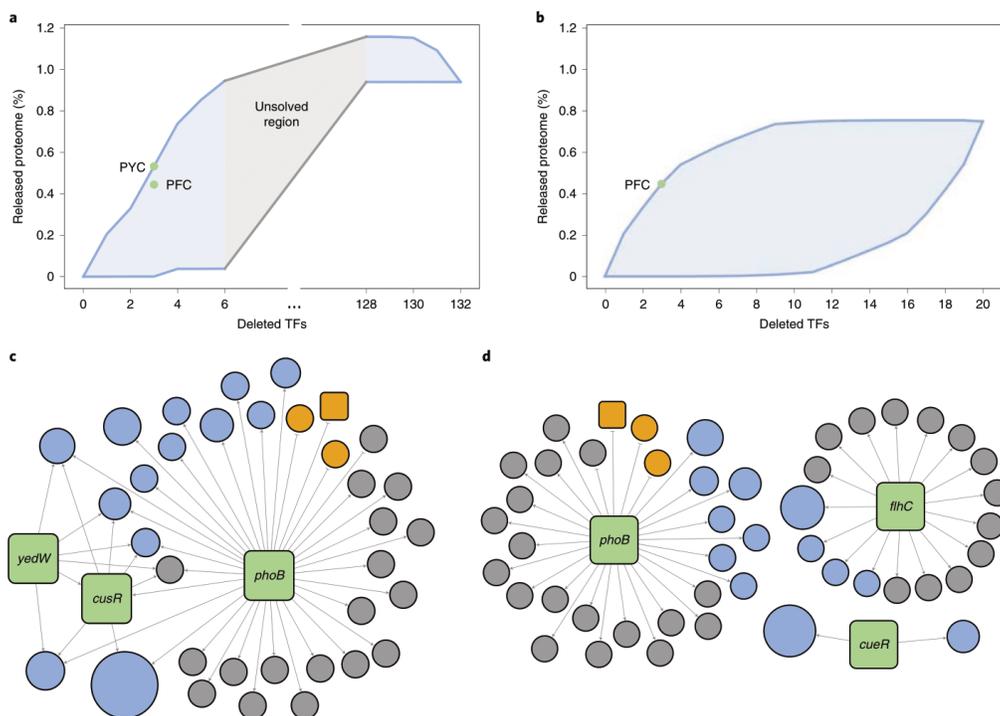


Fig. 3 | Proteome liberation calculations using ReProMin. a, Potential proteome liberation landscape corresponding to the glucose shared-target case; the solved region is shown in blue while the unsolved region in gray. **b**, Solved proteome liberation landscape for the unique-target case in glucose. The locations in the landscapes of the generated mutants are shown with a green circle. Subnetwork of predicted regulated targets of the best three KO **c,d**, Shared (**c**) and unique (**d**) target cases. In both cases, green squares represent deleted TFs, blue circles predicted silenced targets, yellow circles predicted induced targets and gray circles genes with no proteomic coverage; the size of the circles is proportional to the P_i of the target. High-resolution labeled versions of the subnetworks are available in Extended Data Fig. 3.

interaction network and quantitative proteomic data to generate and solve the proteome reduction landscape for a particular TF list, finally returning the n -size combination of mutations that silences the higher proteomic load in a particular growth condition (see Methods). Using ReProMin we computed some parts of a proteome liberation landscape depending on the number of combinations (Fig. 3a). For a particular TF list, the proteome liberation landscape boundaries are defined by the TF combination that releases the highest and lowest P_i for all the possible TF combination sizes; meaning that the larger the TF list, the larger and complex the landscape space. The computing time required to solve a landscape is defined by the number of possible combinations that increments exponentially as the TF list becomes larger (see equation (2) in Methods)

ReProMin calculations reveal potential proteome release. We used the glucose minimal medium condition as the starting point for the analysis, since we defined the gene and TF essentiality in that condition from experimental data. We considered two cases to do the calculations. The shared target case: nonessential TFs with positive P_i (132 TFs), which takes into account some TFs with no unique regulated genes and the unique target case: considering previously defined candidate TFs with a positive P_i (20 TFs).

For the shared target case the computational tool was able to solve combinations up to six TFs ($>6 \times 10^9$ combinations roughly taking 95 h of computation, Supplementary Table 3), but we did not evaluate the next case due to its extremely large number of combinations (seven TFs $>11 \times 10^{10}$ combinations). To maximize the solved landscape space, we continued resolving other areas of the space where the number of combinations is small enough to be solved (when $n \rightarrow 0$ or $n \rightarrow r$, see Methods, equation (2)).

Our calculations revealed that the elimination of all 132 nonessential TFs is not the best case of proteome liberation. However, the elimination of 130 TFs would potentially liberate up to 1.06% of the total *E. coli* proteome but up to 0.94% can be liberated by removing the top combination of six TFs, while up to 0.53% of the total proteome can be released by removing a top combination of three TFs (Fig. 3a).

For the unique target case, the elimination of the entire 20 candidate TFs would liberate up to 0.72% of the proteome, and our simulations show that there is not a significant improvement in resource release after the elimination of eight TFs. In fact, 60% of the total potential liberation can be achieved by removing just the top three TFs (Fig. 3b).

We also performed ReProMin calculations for other conditions in which proteomic data is available, such as growth on galactose,

acetate or glycerol + casamino acids (casAA) minimal medium and rich LB medium. In each case, we used the specific set of essential genes (see Methods) and for the environment specific genes we ran essentiality simulations with a genome-scale metabolic model in the corresponding growth condition (see Methods).

As a result, we obtained 164 nonessential TFs for galactose, 166 for acetate, 171 for glycerol + casAA and 165 for rich LB medium. We found that most of the identified dispensable TFs (89%) are shared among all the tested conditions. We identified 22 candidate TFs for galactose and acetate that belong to the unique target case and have a positive P_b . Proteome liberation calculations were made using these subsets of candidate TFs, predictions show that we can release 0.88 and 0.81% of the total proteome in galactose and acetate, respectively, with the deletion of all these TFs (Extended Data Fig. 2a,b). For glycerol, we found 24 candidate TF, liberating up to 2.9% of total proteome (Extended Data Fig. 2c) and finally for rich LB medium we identified 20 candidate TF, with a potential 0.5% of liberation, being the worst condition for potential proteome liberation tested (Extended Data Fig. 2d). This is in agreement with previous proteome usage analysis, where it has been showed that at higher growth rates there is less of the dispensable proteome^{10,18}.

Generation of combinatorial strains. Even though *E. coli* is one of the most studied organisms and its TRN has been widely investigated, only half of its genes have regulatory information (RegulonDB). We prevented detrimental effects on gene expression due to our incomplete knowledge of the regulatory network by selecting the smallest combination of TFs that liberates the greatest amount of resources.

Based on our ReProMin predictions, we generated two triple knockout (KO) strains for the best combinations of the shared target and unique target cases predicted in glucose minimal medium. In the shared target case, this corresponded to the strain PYC (Δ *phoB* – phosphate scavenging system, Δ *yedW* – unknown gene, Δ *cusR* – copper/silver export system regulator) with a P_b of 1.3 fg representing 0.53% of the total proteome in glucose growth conditions (Fig. 3a). This design was a particular case of shared regulation where most of the target genes are only silenced by the deletion of all the three TFs together (Fig. 3c). In the unique target case the resulting strain was PFC (Δ *phoB*, phosphate scavenging system; Δ *flhC*, flagella master regulator and Δ *cueR*, copper efflux system) with a P_b of 1.08 fg representing 0.44% of the total proteome in glucose (Fig. 3b). The unique target case has a higher degree of confidence in the design than the shared target case due to a simpler regulatory subnetwork, since we noticed that unique target cases mainly belong to the single input module network motif (SIM)¹⁷ (Fig. 3d). We also generated a strain, using an intuitive approach, in which we eliminated three TFs that regulate nongrowth-related functions. To construct this strain, we randomly selected three TFs from those previously found downregulated in strains with regulatory mutations selected by adaptive laboratory evolution (ALE)⁹. The resulting strain was called FOG (Δ *flhA*, *oxyR*, *gadE*) the genes code for flagella sigma factor, oxidative stress master regulator and acid resistance regulator, respectively. The FOG strain was not generated by our design pipeline; therefore, the regulatory interventions affect some essential functions (nine essential genes) mainly involved in de novo synthesis of AMP and heme groups and it was used as a control for comparison with the computationally designed strains.

Finally, to test ReProMin in a different environment for which high confidence gene essentiality data is available, we also constructed the best 3KO strain in LB rich medium (Extended Data Fig. 2d). The resulting strain was PYN (Δ *phoB*, phosphate scavenging system; Δ *yqhC*, furfural reduction¹⁹ and Δ *ntrC*, nitrogen regulation two-component system) potentially liberating 0.33% of the total proteome.

RNA-seq confirmed the specificity of introduced mutations.

The predictive power of ReProMin depends on the accuracy of the interactions captured in the *E. coli* TRN. We validated the predicted transcriptional changes for the case PFC that showed the greatest proteome release. We measured PFC and the wildtype (WT) strains' transcriptome profiles obtained by RNA-seq in the same environment. This experiment aimed at determining the degree of success in gene silencing at the transcriptional level, and at assessing other possible transcriptional perturbations resulting from our regulatory modifications. Our results showed that no transcripts corresponding to the three deleted TFs were detected in PFC (Extended Data Fig. 4a). By mapping the fold change obtained in the analysis to the predictions of the computational tool, it was possible to visualize the impact at the transcriptional level of the missing regulators on their targets (Extended Data Fig. 4c). Four targets associated to *flhC*, corresponding to genes forming the flagella (*flgB*, *flgC*, *flgE* and *flgG*) were completely silenced; furthermore, all the other flagella-related genes also registered a decrease on their expression. Regarding *phoB*, two targets were successfully silenced (*phnI* and *phnL*), both genes belong to an operon that is induced under phosphate starvation and is required for use of phosphonate and phosphite as phosphorous sources²⁰, many other targets related to this operon also reduced their expression. On the contrary, *phnK* present in the same operon was overexpressed. We were unable to map any transcripts to six genes belonging to the previously mentioned operon, which may not be entirely expressed in the absence of phosphate starvation. Furthermore, *phoR* (part of the *phoB*–*phoR* two-component system) also reduced its expression. Finally, both targets of *cueR* (*copA* and *cueO*) also decreased drastically their abundance.

Regarding the accuracy of our ReProMin predictions, 30 genes of 47 predicted silenced genes were silenced at different levels, whereas nine predicted silenced genes presented higher expression values than the WT strain. Carefully reviewing discrepancies among predicted and measured transcriptional changes revealed that some evidences on which the *E. coli* TRNs are built, are computationally predicted or derived from indirect observations. These weak pieces of evidence resulted in false predictions and may be refined if those networks are improved. Finally, transcripts of eight predicted targets were not found in either strain. These observations show that in 72% (28 of 39 measured genes) of the cases the predictions of the computational tool were accurate (Extended Data Fig. 4d).

In addition to the designed transcriptional changes, we found 17 genes differentially expressed (eight downregulated genes and nine upregulated) (Extended Data Fig. 4b and Supplementary Table 4). This RNA-seq analysis shows that besides the intended transcriptional changes, few off-target effects were identified at the transcriptomic level in the PFC strain growing on glucose M9 medium.

To theoretically quantify the actual mass of the liberated proteome fraction, we did an estimation of the translation rates of the gene targets associated to the deleted TFs plus the previously mentioned differentially expressed genes (Fig. 4a) (see Methods).

The calculations showed that the mass of the proteome fraction associated to the removed TFs targets decreased from a 0.48% of the total proteome in the WT to a 0.25% in PFC, CueR being the TF that contributes most to the liberation (Fig. 4b). Initially, ReProMin calculations estimated a liberation of 0.44% of the total proteome in PFC, according to these calculations (with the uncertainty that they may have) this would correspond to a 57% of success in the estimated resource liberation. Liberated resources are presumably redistributed for upregulated genes, which show a larger proteome fraction in the mutant (Fig. 4c), and could become available for the expression of engineered functions.

UT designed strains show increased cellular budget. Our three experimentally generated mutants (shared target case, unique target case and control) were evaluated in rich (LB) as well as in minimal

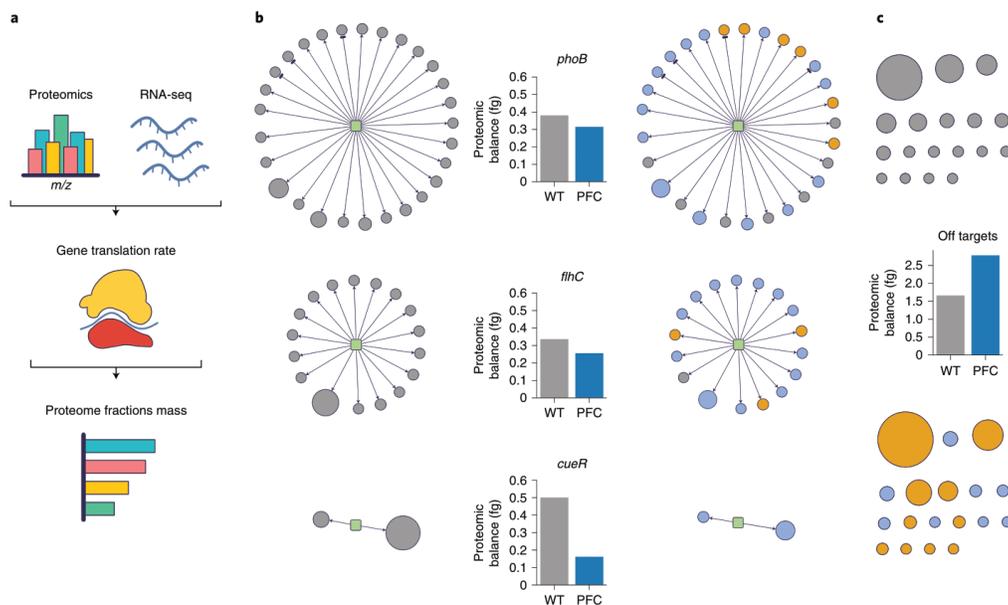


Fig. 4 | Deleted TF proteome fraction mass estimation. **a**, Pipeline used to estimate the proteome changes from RNA-seq data. **b**, Proteome fraction mass representation and comparison for the three deleted TFs associated targets for WT (left) and PFC (right). **c**, Proteome fraction mass representation and comparison of the identified off-targets for WT (top) and PFC (bottom). Green squares represent deleted TFs; in the case of PFC, blue circles signify targets that reduced their mass compared to the WT, yellow circles targets that increased their mass and gray circles targets not found expressed thus not contributing to the mass; the size of the circles is proportional to the P_i of the target. In all cases, graph bars represent the sum of the mass of all genes in each group.

medium containing three different carbon sources (acetate, galactose, glucose). The ReProMin designed strains (PYC and PFC) showed neither growth defects nor increase in the growth rate or biomass yield in any of the four conditions tested. On the contrary, the control strain (FOG) showed growth defects in all growth conditions tested (Extended Data Fig. 5). This phenotypic defect shown by FOG strain may be the result of the elimination of the principal acid resistance system activator (*gadE*), since we are using batch growth with no pH control. For glucose minimal medium, we also evaluated the effect of recombinant protein production using a plasmid expressing a genetic circuit with two fluorescent reporters (Fig. 5a)²¹. The burden caused by carrying a plasmid was reflected as a decrease in the growth rate in all tested strains (Extended Data Fig. 6a); this decrease is higher when the plasmid was expressing the genetic circuit; however, the burden displayed by both ReProMin designed strains was lower compared to the WT counterpart. Additionally, the PFC strain also showed a higher final biomass production (Extended Data Fig. 6b). It has been described that the expression levels of two protein reporters encoded on the same plasmid but without a regulatory connection between them is captured by a linear relationship that can be interpreted as an isocost line, a concept used in microeconomics to describe how two products can be bought with a limited budget, so the more is used on one, the less can be used on the other. These lines represent the boundary of the production budget of a given strain and growth condition (Fig. 5b)²¹. We obtained the isocost lines at balanced growth, defined as growth during the exponential phase in which a steady-state of

the cellular concentration of both green (GFP) and red fluorescent protein (RFP) reporters is achieved (in other words, the cellular concentration of fluorescent protein does not change in time after roughly 5 h) determined by two different methods: mean plate reader fluorescence and mean fluorescence measured by flow cytometry. The line corresponding to PFC strain shows a parallel upward shift compared to the WT strain, which represents an increase of 9% in absolute fluorescence (Fig. 5c), 5% in normalized fluorescence per cell (Fig. 5d) and 12% in mean fluorescence per cell (Extended Data Fig. 7). This difference is increased at the stationary phase of the culture (~22 h) where higher maximal biomass is achieved and the quantity of recombinant protein is increased up to 18% (Extended Data Fig. 9a). The shared target case strain (PYC) showed no budget increase compared to the WT strain in absolute fluorescence (Fig. 5c), however it showed increased budget in normalized fluorescence (Fig. 5d). The PYC mutant show reduced maximal biomass production (Extended Data Fig. 6b), these results suggest that there is increased budget but also reduced growth capacity, the latter probably due to unexpected regulatory interactions resulting from the more complex case of the shared target case.

Our ME-model simulations showed that reducing the unmodeled protein fraction (UPF) from 0.36 to 0.25, increase the maximum recombinant proteome sector from 0.15 to 0.20 (Extended Data Fig. 1). These increases in recombinant protein production are also in agreement with the experimental observations in previous works²², where for glucose minimal medium growth conditions the maximum observed heterologous fraction was 14% of the total

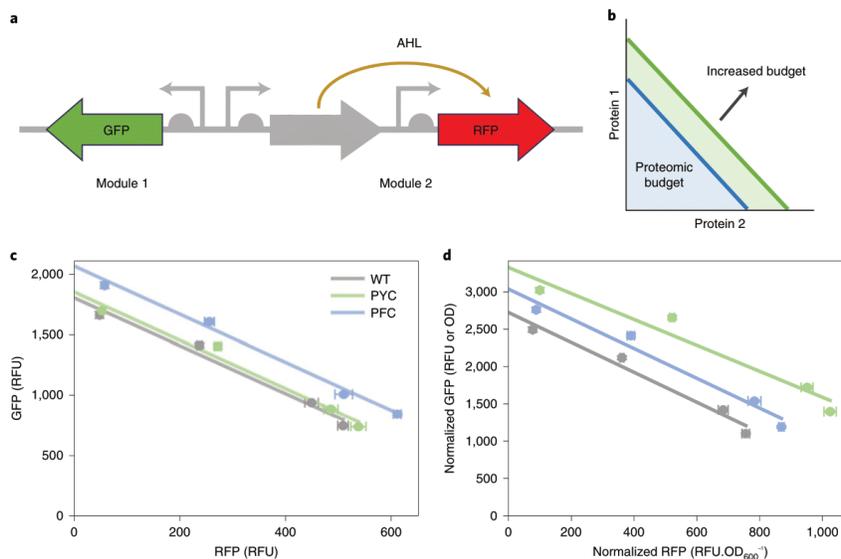


Fig. 5 | Synthetic circuit characterization. **a**, Schematic of the gene circuit evaluated, which encodes the fluorescent reporters GFP and RFP: the first is constitutively expressed, while the latter is under the control of a AHL-inducible promoter. **b**, When plotting the expression of one protein against the other at different levels of induction, an isocost line is obtained. The size of the area below the line represents the total proteome budget dedicated to the circuit, an upward shift in the line represents an increase in the budget. **c,d**, Isocost lines of the designed strains during balanced growth showing absolute fluorescence (**c**) and normalized fluorescence (**d**). In all cases, points represent the Gaussian fitted fluorescence value ± 2 s.d. for $n=9$ (see Methods) of red reporter (x axis) plotted against the green reporter (y axis) in an increasing inducer concentration (0, 2.5, 5, 20 nM AHL). A linear regression was used to fit the points to a line.

proteome, then a 0.5–1% increase in proteome availability may result in a 10% increase in the heterologous budget, which is close to what we found in our PFC strain.

We also evaluated and constructed the fourth best combination of mutants for glucose minimal medium. ReProMin predicts as PFC with the addition of *marA* as the best fourth TF to KO in glucose minimal medium, with a 0.54% of P_B . The resulting strain, PFCM ($\Delta marA$, multiple antibiotic resistance), also has a higher proteomic budget than the WT, however, it is not higher than PFC at balanced growth (Extended Data Fig. 8a), confirming our hypothesis that lower number of genetic interventions will have a higher probability of success due to epistatic interactions of multiple TF deletions. Additionally, to experimentally test ReProMin predictions in a different environment we evaluated the previously described PYN mutant in rich LB medium. PYN has a higher proteome budget than the WT strain during balanced growth on LB, and also at stationary phase (Extended Data Figs. 8b and 9b) that demonstrates that our method can be used in any condition where gene essentiality and proteomics information is available.

Expression of a heterologous metabolic pathway. We tested the ability of our engineered strain for synthesizing the molecule violacein as a proof-of-concept for applications of the mutants designed by our method, with a higher proteomic budget, in metabolic engineering using a costly heterologous pathway. Violacein is a pigment from *Chromobacterium violaceum* endowed with many biological activities (antibacterial, antiviral, antiparasite) and has recently gained importance in the industrial field especially for applications in cosmetics, medicines and fabrics²³. Violacein is synthesized in a

five-step metabolic pathway using tryptophan as a precursor. Here, we used the violacein pathway plasmid reported by Darlington et al.¹³, where the five genes for violacein biosynthesis are arranged in two operons, one consisting of *vioA* constitutively expressed, while the rest of the pathway encoded by the *vioBCDE* genes is under the control of an *N*-acyl-homoserine lactone- (AHL-) inducible promoter¹³ (Fig. 6a). This construction follows the same principle as the previous circuit so the more of one module is produced, the less of the other is expressed due to the competition for limited resources for gene expression. However, in this case the number of genes in each module is different and code for actively metabolic enzymes with different kinetic properties, which results in differential violacein biosynthesis. This system is ideal to test the ReProMin designed strains, since their capacity to produce a metabolite from a heterologous pathway is directly dependent on the fraction of the cellular machinery that can be allocated for the expression of accessory proteome in a range of conditions (from low to high competition).

We evaluated violacein production after 24 h in the WT and PFC strains using M9 glucose medium with and without tryptophan (2.0 g l^{-1}), in both cases AHL (1.25, 2.5, 5, 10, 20 nM) was added for induction. We found that the maximum production was achieved with just a minimum amount of inducer (1.25 nM) indicating that it is crucial to have a balanced expression of the pathway with the right amount of each module to maximize the synthesis of the final product. PFC showed a mean increase in mg of violacein production of 18% ($P < 0.05$) (Fig. 6b) and 20% in normalized data (mg violacein per mg protein) (Extended Data Fig. 10a). Even though the designed regulatory interventions do not target the metabolic network of the organism, we still found a significant increase in

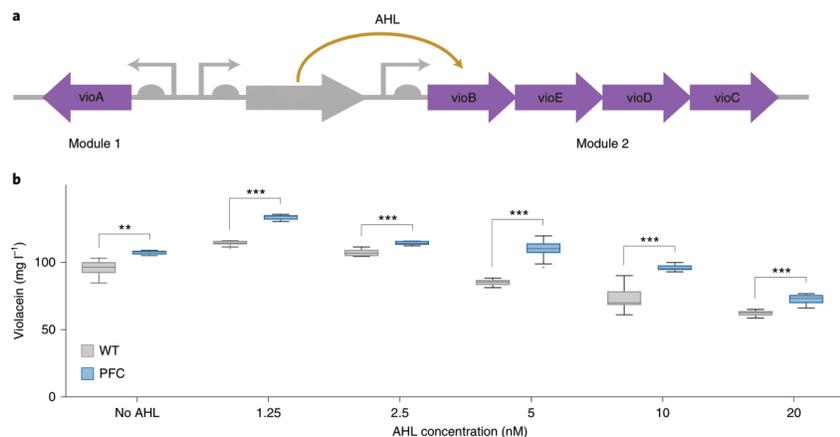


Fig. 6 | Violacein pathway evaluation. **a**, Schematic of the circuit expressing the pathway for violacein biosynthesis. **b**, Total violacein production using 2 g l^{-1} tryptophan after 24 h in the presence of increasing inducer (AHL) concentrations (mean \pm s.d., $n=9$). Asterisks *, ** and *** denote significant differences between WT and PFC using a two-tailed unpaired Student's *t*-test. The following *P* values were obtained for AHL concentrations: No AHL, $P=0.0031$; 1.25 nM, $P<0.0001$; 2.5 nM, $P<0.0001$; 5 nM, $P<0.0001$; 10 nM, $P<0.0001$; 20 nM, $P<0.0001$.

violacein production, even without tryptophan addition in a non-metabolically engineered background and without further culture condition optimization (Extended Data Fig. 10b). Previous studies have shown that proteome reallocation, such as the obtained by ALE do not require metabolic flux distribution²⁴; therefore, we do not expect a relevant flux redistribution in PFC strain to be responsible for the increase in violacein production, although it would be interesting to measure it. This increase, presumably resulting from a better expression of the heterologous pathway, shows that our approach can also be harnessed to increase the production of metabolites from costly heterologous metabolic pathways.

Discussion

Gene regulatory networks are robust and can be severely rewired with interesting phenotypic outcomes²⁵, thus they are a perfect rational engineering target for synthetic biology applications. In this work, we have proved that the definition of an essential gene set together with regulatory network information allows the identification of TFs whose elimination leads directly to silencing proteome fractions that are not used in a particular condition. We have showed that by eliminating hedging proteome activators we can release resources and increase cellular capacity for engineered functions. The certainty of ReProMin calculations will always depend on gene essentiality data and the knowledge of gene-TF interactions. In this case, due to the gaps in our understanding of the gene-TF interactions of *E. coli*, we found that it is better to use the unique target cases where the combinations are less complex.

In agreement with the presented metabolism and gene-expression-model simulations reducing the unused protein fraction, our designed strain showed a higher proteomic budget, measured by the isocost lines, and a higher capacity to produce a metabolite from a heterologous pathway. Even though the amount of reduced proteome in the PFC strain may seem insignificant in the evaluated conditions, we have shown that a 1% reduction of unused proteome can increase the heterologous proteome fraction by 10%. The total amount of proteome available for reduction is dependent on the growth conditions and different methods of calculation may yield different numbers. For example, O'Brien et al. showed that

in glucose minimal medium up to 30% of the proteome may be unused¹⁰. According to our classification of essential gene list (that includes several experimental datasets) and comparing metabolism and gene-expression-model proteome use predictions to actual proteomic data¹⁶, we calculate that a 33.6% of proteome is directly available for reduction. Using our method, we were able to liberate a 1.3% in the best of the analyzed cases. Most of the genes that contribute to the dispensable proteome have no known TF or are regulated by a TF classified as essential, therefore are out of the scope of ReProMin. Furthermore, analyzing the data of a genome reduced strain of *E. coli*²⁶, we found that a genome reduction of 15% (743 genes) led to a proteome reduction (according to our calculations with the same proteomic dataset) of only 1.5% in the same condition (glucose minimal medium). Therefore, reducing the full 30% of dispensable proteome remains a challenge that may be tackled using a combination of approaches, such as targeting global regulators, reducing large sections of the genome or by the optimization of the core proteome²⁷.

By comparing our strains with an intuitive control strain, we show that inaccurate TF elimination results in detrimental effects on growth, maximal biomass and protein production (Extended Data Figs. 6 and 9). These findings indicate that the elimination of a combination of TFs is not a trivial process; it may affect essential functions and introduce phenotypic defects. Our method shows good accuracy in terms of the obtained gene expression changes measured by RNA-seq, despite our limited knowledge of the regulatory networks. In addition, the regulatory data available is condition dependent, which limits the predictive power of our method, since we need to assume that regulatory interactions are present at all times. We anticipate that developments in high-throughput technologies (such as chromatin immunoprecipitation-seq) combined with new computational approaches²⁸⁻³⁰ will enable the fast generation of complete regulatory networks, that combined with absolute quantification of proteomes or translation rates, will enable the application of ReProMin method to even nonmodel organisms.

We presume liberated resources are redistributed mainly among the detected upregulated genes that proteome fraction is bigger in the mutant (Fig. 4c), and we expect these resources to be redirected

to the expression of engineered functions when introduced to the modified strains.

Several experimental approaches have been applied for resource allocation optimization in bacterial host engineering. Genome minimization has been mainly done by large scale genetic interventions whose outcomes are difficult to predict and do not show greater genome stability^{31,32}. ALE has showed great success, especially to identify functions not related to growth³³; however, it selects for fast growing strains that do not necessarily result in the best production phenotypes. Moreover, the underlying selection mechanisms in ALE are normally not known therefore its effects are not predictable³⁴. Genome-scale models, such as the metabolism and gene-expression model, may also be used to find the proteomic cost and fitness benefit of gene expression, thus aiding in the design of proteome allocation; however, kinetic data of each protein is needed³⁵ and its scope focuses on growth related functions. There are only a few reports describing regulatory approaches to improve production phenotypes, such as the global transcriptional machinery engineering³⁶, but none of them followed a rational approach. The methodology presented in this work is a new strategy for proteome optimization with minimal genetic interventions that overcomes the serious limitations of deleting large regions of the genome; it is a flexible pipeline that can be applied to other growth and production conditions and also to different organisms where sufficient information is available. This work shows the potential of rational design of biological systems over the predominantly used trial and error approaches.

Online content

Any Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-020-0593-y>.

Received: 30 August 2019; Accepted: 15 June 2020;

Published online: 13 July 2020

References

- Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253–aad6253 (2016).
- Balikó, G. et al. in *Synthetic Biology: Parts, Devices and Applications* (eds Smolke, C., Lee, S. Y., Nielsen, J. & Stephanopoulos, G.) 49–80 (Wiley-VCH, 2018).
- Ishihama, A. Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.* **34**, 628–645 (2010).
- Ulrich, L. E. & Zhulin, I. B. The MIST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res.* **38**, D401–D407 (2010).
- Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).
- Isalan, M. et al. Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840–845 (2008).
- Mitchell, A. et al. Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220–224 (2009).
- Tagkopoulos, I., Liu, Y.-C. & Tavazoie, S. Predictive behavior within microbial genetic networks. *Science* **320**, 1313–1317 (2008).
- Utrilla, J. et al. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. *Cell Syst.* **2**, 260–271 (2016).
- O'Brien, E. J., Utrilla, J. & Palsson, B. O. Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. *PLoS Comput. Biol.* **12**, e1004998 (2016).
- Nikolados, E.-M., Weiß, A. Y., Ceroni, F. & Oyarzún, D. A. Growth defects and loss-of-function in synthetic gene circuits. *ACS Synth. Biol.* **8**, 1231–1240 (2018).
- Ceroni, F. et al. Burden-driven feedback control of gene expression. *Nat. Methods* **15**, 387–393 (2018).
- Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, 695 (2018).
- Lloyd, C. J. et al. COBRAME: a computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput. Biol.* **14**, e1006302 (2018).
- Santos-Zavaleta, A. et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).
- Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2015).
- Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
- Kim, J., Darlington, A., Salvador, M., Utrilla, J. & Jiménez, J. I. Trade-offs between gene expression, growth and phenotypic diversity in microbial populations. *Curr. Opin. Biotechnol.* **62**, 29–37 (2020).
- Turner, P. C. et al. YqhC regulates transcription of the adjacent *Escherichia coli* genes yqhD and dlkA that are involved in furfural tolerance. *J. Ind. Microbiol. Biotechnol.* **38**, 431–439 (2011).
- Yakovleva, G. M., Kim, S. K. & Wanner, B. L. Phosphate-independent expression of the carbon-phosphorus lyase activity of *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **49**, 573–578 (1998).
- Gyorgy, A. et al. Isocost lines describe the cellular economy of genetic circuits. *Biophys. J.* **109**, 639–646 (2015).
- Bienick, M. S., Young, K. W., Klesmith, J. R., Detwiler, E. E. & Tomek, K. J. The interrelationship between promoter strength, gene expression, and growth rate. *PLoS ONE* **9**, 109105 (2014).
- Durán, N. et al. Advances in *Chromobacterium violaceum* and properties of violacein-Its main secondary metabolite: a review. *Biotechnol. Adv.* **34**, 1030–1045 (2016).
- Long, C. P., Gonzalez, J. E., Feist, A. M., Palsson, B. O. & Antoniewicz, M. R. Fast growth phenotype of *E. coli* K-12 from adaptive laboratory evolution does not require intracellular flux rewiring. *Metab. Eng.* **44**, 100–107 (2017).
- Baumstark, R. et al. The propagation of perturbations in rewired bacterial gene networks. *Nat. Commun.* **6**, 10105 (2015).
- Posfai, G. et al. Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–1046 (2006).
- Hidalgo, D. & Utrilla, J. in *Minimal Cells: Design, Construction, Biotechnological Applications* (eds Lara, A. & Gosset, G.) 211–230 (Springer International Publishing, 2020); https://doi.org/10.1007/978-3-030-31897-0_8
- Fang, X. et al. Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proc. Natl Acad. Sci. USA* **114**, 10286–10291 (2017).
- Ibarra-Arellano, M. A., Campos-González, A. I., Treviño-Quintanilla, L. G., Tauch, A. & Freyre-González, J. A. Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria. *Database* **2016**, baw089 (2016).
- Sastry, A. V. et al. The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **10**, 1–14 (2019).
- Couto, J. M., McGarrity, A., Russell, J. & Sloan, W. T. The effect of metabolic stress on genome stability of a synthetic biology chassis *Escherichia coli* K12 strain. *Microb. Cell Fact.* **17**, 8 (2018).
- Choe, D. et al. Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* **10**, 935 (2019).
- Yang, L. et al. Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Sci. Rep.* **6**, 36734 (2016).
- McCloskey, D. et al. Evolution of gene knockout strains of *E. coli* reveal regulatory architectures governed by metabolism. *Nat. Commun.* **9**, 3796 (2018).
- Nilsson, A., Nielsen, J. & Palsson, B. O. Metabolic models of protein allocation call for the kinome. *Cell Syst.* **5**, 538–541 (2017).
- Klein-Marcuschamer, D., Santos, C. N. S., Yu, H. & Stephanopoulos, G. Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. *Appl. Environ. Microbiol.* **75**, 2705–2711 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Metabolism and gene-expression-model simulations. All simulations were done using the model iJL1678b-ME¹⁴. The corresponding transcription and translation reactions for recombinant protein (GFP) production were manually added to the model using standard methods. Unused protein fraction and flux through the recombinant protein production are changeable variables in the metabolism and gene-expression model that affect predicted growth rate and proteome composition, the values of these two variables were systematically changed in the metabolism and gene-expression model to assess their effect on growth rate (UPF = 0.36, 0.30, 0.25; flux = 0, 0.001, 0.002, 0.0025, 0.0030, 0.0035, 0.0040), all other model parameters were set as default. Proteome sectors were classified according to O'Brien et al.¹⁰.

Definition of the essential gene list. To compile the essential gene list in the glucose minimal medium condition we combined five different datasets from different sources. Three of them were experimentally generated using different methods of gene disruption: (1) random transposon mutagenesis using M9 with glucose as growth condition (Tn-seq)¹⁷, (2) removing large fragments of the chromosome using a homologous recombination system in rich medium (LB)¹⁸ and (3) the updated list of the mutants of the Keio collection that are lethal, the collection was generated using rich medium^{19,40}. Two gene lists were generated in silico using simulations of genome-scale metabolic and expression models capable of predicting gene expression needs in a particular condition: (4) genes that are essential for growth in M9 with glucose using iOL1554-ME model¹⁴ and (5) genes that are essential for growth in the metabolic model iJO1366 and also experimentally in M9 with glucose¹⁴. Within the compiled list, genes exclusively belonging to the Tn-seq and the glucose minimal medium metabolism and gene-expression-model simulation gene lists were considered conditionally essential, as these gene lists were originally generated using M9 with glucose as the growth condition, while the rest of the genes were classified as core essential for our purposes. For the cases of galactose, acetate and glycerol + casAA minimal medium conditions, we performed gene essentially analysis with the iML1515 metabolic model¹⁵ in COBRAPy⁴¹.

For the rich LB medium case, we used the Keio mutants list.

Identification of candidate regulators and combinatorial analysis. We sorted the TF-gene interactions from RegulonDB (v.8, regulondb.ccg.unam.mx), discarding all the sigma factor-gene interactions. Next, we classified as essential all TFs that activate at least one essential gene (from the condition-specific essential gene list) and as nonessential all TFs that do not activate any essential genes. Then we analyzed the subnetwork of interactions of each nonessential TF by numerically analyzing the output level of each TF (TF_{OUT}), which is classified into positive and negative output (TF_{OUT+} and TF_{OUT-}) representing positive and negative regulated genes, respectively, and the degree of entry of each regulated gene (GENE_{IN}) in turn also divided into positive and negative (GENE_{IN+} and GENE_{IN-}). We defined as candidate for proteome reduction all those TFs that activate at least one unique gene, which numerically meets the following condition:

$$TF_{OUT+} \geq 1 \wedge GENE_{IN+} = 1 \wedge GENE_{IN-} \geq 0 \quad (1)$$

where GENE_{IN} represents any gene activated by TF.

On the other hand, the proteomic dataset previously described¹⁷ was used to calculate the P_i of each gene and P_o of each TF according to the equations in Fig. 2.

ReProMin combinatorial analysis was achieved as follows, given a list of TFs, the total number of combinations was calculated with this formula:

$$P(n, r) = \frac{n!}{r!(n-r)!} \quad (2)$$

where n represents the number of candidate TFs, and r the size of the combination ($r \leq n$).

Next, the total number of silenced and induced genes for each combination was determined following the next criteria: for every gene involved in the combination, we subtracted one from the value of GENE_{IN+} for each TF that regulates the target gene positively and one to the value of GENE_{IN-} for each TF that regulates the target gene negatively. At the end of this process a gene was considered silenced if:

$$GENE_{IN} = GENE_{IN+} = 0 \quad (3)$$

or induced if:

$$GENE_{IN} = GENE_{IN-} = 0 \quad (4)$$

Finally, the P_o of each combination tested was calculated and ranked.

The full computational set of tools coded in Python and datasets used in the analysis are available in the following repository (<https://github.com/uttrillalab/repromin>). Cytoscape software v.3.7 (ref. ⁴²) was used to plot the network representation of the data.

Generation of combinatorial knockout strains. The combinatorial mutants were generated by sequential P1 phage transduction from the individual knockout

strains of the Keio collection according to the protocol described by Miller⁴⁶. The removal of the kanamycin resistance cassette before each transduction was done using the pE-FLP plasmid (Addgene plasmid no. 45978), pE-FLP was a gift from D. Endy and K. Shearwinand⁴⁷. Each knockout strain was confirmed by PCR using primers flanking each gene. In all experiments *E. coli* BW25113 was used as the WT background. The characteristics of the strains, plasmids and primers used in this study are described in the supplementary material (Supplementary Table 5).

RNA sample extraction and sequencing. Strains were grown in 50 ml of M9 medium with glucose (4 g l⁻¹) M9 medium in 250 ml Erlenmeyer flasks cultures in an orbital incubator at 37 °C (250 r.p.m.). Cells were collected in mid-log phase using the Qiagen's RNeasy Protect Bacteria reagent according to the manufacturer's specifications. Cell pellets were incubated with lysozyme, SuperaseIn and protease K for 10 min at 37 °C. Total RNA was isolated and purified using Zymo Research's Quick-RNA kit according to the manufacturer's specifications. All samples' quality was inspected in a bioanalyzer RNA chip (Agilent). Starting with 10 µg of total RNA of each sample, the removal of ribosomal RNA was done with the Ribominus kit by Invitrogen. For the construction of the libraries, the TruSeq Stranded mRNA HT Sample Prep Kit by Illumina was used. For sequencing a NextSeq 500 v.2 was used, with a configuration of 2 × 75 paired-end read and 10 million reads per sample.

Reads were mapped to reference genome *E. coli* MG1655 (RefSeq, NC_000913.3) using aligner Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2>). Final differential analysis was made using the Cufflinks library (<http://cole-trapnell-lab.github.io/cufflinks>). Genes with a log₂ fold change that was ≥ 1 were considered upregulated and ≤ -1 were considered downregulated, considering $P \leq 0.01$ and $n = 2$.

Estimation of the theoretical proteome from RNA-seq data. To do an estimation of the translation rates from proteomics data and from transcript abundances, we assumed that the transcripts in our RNA-seq data yields the proteome reported by Schmidt et al. in glucose, so we calculated each gene translation efficiency rate (si) using this equation:

$$si = \frac{C_{cell}^i}{ri}$$

where ri represents the raw fragments per kilobase of transcript per million mapped reads value of each gene and C_{cell}^i the number of protein copies reported. For all the genes with no proteomic data we assumed a fixed rate corresponding to the mean rate of the reported genes. Then, we used these rates to estimate the number of protein copies per cell (Pi) for our mutant PFC according to:

$$Pi = ri \times si$$

Finally, the gene load was calculated as previously described.

Growth phenotype characterization. For the evaluation of growth in different carbon sources, the following conditions were used: glucose M9 medium (4 g l⁻¹), galactose M9 medium (3.2 g l⁻¹), acetate M9 medium (2.5 g l⁻¹) and LB rich medium. Cells were cultured overnight in the corresponding medium. The next day the strains were diluted to an optical density (OD₆₀₀) of 0.05 in fresh medium and 150 µl of the fresh culture was transferred to a transparent 96-well plate (Corning) and incubated at 37 °C with fast linear shaking in a micro-plate reader (Synergy 2.0, BioTek) for 24 h, taking measurements for OD₆₀₀ every 20 min. The characterization of the growth kinetics was conducted using the algorithm Fiterider (v.1.0) developed by Swain et al.⁴⁸ with the default parameters. The algorithm performs a Gaussian fit of the raw data and in all cases the resulted fitted value ± 2 s.d. (ensuring a confidence of at least 95% ($P \leq 0.05$)) was used when comparing the mutant strains against the WT.

Isocost circuit evaluation. Strains were inoculated into glucose M9 medium with gentamicin (20 µg ml⁻¹), and grown overnight. Next day, strains were diluted to an OD₆₀₀ of 0.05 in fresh glucose M9 medium containing AHL (Sigma-Aldrich, final concentrations of 1.25, 2.5, 5, 10, 20 nM), then 150 µl of the fresh culture was transferred to a 96-well black plate with transparent bottom (Corning) and incubated as described previously, taking measurements for OD₆₀₀, GFP (excitation, 485 nm and emission, 528 nm) and RFP (excitation, 590 nm and emission, 645 nm). The characterization of the production kinetics of GFP and RFP was also done using the algorithm described above.

Flow cytometry measurements. For flow cytometry measurements, cell cultures were prepared as described before, but later grown in 24-well plates using 1 ml of medium. Every hour, 50 µl aliquots were taken from each well and mixed with 150 µl of PBS, the volume of the wells was kept constant by adding fresh medium. Cell suspension was loaded into an Attune NxT Flow Cytometer (ThermoFisher) and analyzed for GFP (excitation, 488 nm and emission, 525/50 nm) and RFP (excitation, 561 nm and emission, 620/15 nm). For each sample 20,000 events were analyzed and population means were estimated using the default software of the instrument. The characterization of the production kinetics of GFP and RFP was also done using the algorithm described above.

Characterization of violacein-producing strains. The strains were inoculated into glucose M9 medium with gentamicin ($20 \mu\text{g ml}^{-1}$), and grown overnight. Next day, strains were diluted to an OD_{600} of 0.05 in fresh glucose M9 medium containing AHL (1.25, 2.5, 5, 10, 20 nM) and tryptophan (2.0 g l^{-1}), then $150 \mu\text{l}$ of the fresh culture was transferred to a 96-well plate and incubated as described before. After 24 h the plate was centrifuged (13,000g, 10 min), and the supernatant of each well was discarded. Violacein was extracted by suspending the pellet in each well in $150 \mu\text{l}$ absolute ethanol and incubating the plate at 95°C for 10 min followed by pelleting cell debris (13,000g, 10 min). Violacein present in the extract was determined spectrophotometrically at 575 nm in a micro-plate reader (Synergy 2.0, BioTek) and quantification was made using a curve with a purchased violacein standard (Sigma-Aldrich).

Quantification of total protein. The Biuret method was used for the quantification of total protein: 1 ml of culture was taken and centrifuged (13,000g, 10 min). The supernatant was collected and the pellet was washed with 0.2 ml of water, resuspended and centrifuged again, the water was discarded. Pellets were resuspended in 0.1 ml of 6 M NaOH and incubated at 95°C for 10 min to break the cells and hydrolyze the proteins. To perform the Biuret reaction, 0.1 ml of 3.2% CuSO_4 was added to the samples and incubated under vigorous agitation for 5 min. Next, samples were centrifuged for 2 min and $150 \mu\text{l}$ of supernatant was placed in a 96-well plate. The absorbance at 555 nm was measured in a plate reader (Synergy 2.0, BioTek).

Statistics. Samples sizes are defined in each figure legend. In all trials, three replicates were included and the experiment was repeated independently on three different days. For growth kinetics and fluorescence, the value of the replicates is presented as the Gaussian fitted value $\pm 2 \text{ s.d.}$ For violacein, results are presented as mean $\pm \text{s.d.}$ and statistical significance between conditions was calculated using Student's *t*-test (two-tailed). All statistical calculations and numerical analyses were performed using Python 3 packages.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

RNA-seq data from this study have been deposited in NCBI's Gene Expression Omnibus (GSE134335). Additional data is available from the corresponding author upon reasonable request.

Code availability

The code and data to run ReProMin can be found at: <https://github.com/utrillalab/repromin>

References

37. Yang, L. et al. Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data. *Proc. Natl Acad. Sci. USA* **112**, 10810–10815 (2015).
38. Kato, J.-I. & Hashimoto, M. Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* **3**, 1–7 (2007).
39. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

40. Yamamoto, N. et al. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* **5**, 335 (2009).
41. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9**, 693 (2013).
42. Orth, J. D. & Palsson, B. Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst. Biol.* **6**, 30 (2012).
43. Monk, J. M. et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
44. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* **7**, 74 (2013).
45. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
46. Miller, J. H. *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* (Cold Spring Harbor Laboratory Press, 1992).
47. St-Pierre, F. et al. One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* **2**, 537–541 (2013).
48. Swain, P. S. et al. Inferring time derivatives including cell growth rates using Gaussian processes. *Nat. Commun.* **7**, 13766 (2016).

Acknowledgements

We thank E. Marquez-Zavala and C. Lloyd for metabolism and gene-expression model simulations support. Y. Castillo-Franco and C. F. Mendez-Cruz for computational support and G. Hernandez-Chavez, H. King, M. Hughes and A. Sicilia for technical support. We acknowledge the funding provided by UNAM-DGAPA-PAPIIT projects IA200716 and IA201518. Newton advanced Fellowship Project NA 160328. J.J. and J.K. acknowledge the support received from the Biology and Biotechnology Research Council (grant nos. BB/M009769/1 and BB/T011289/1) and European Union's Horizon 2020 research and innovation program for the project P4SB (grant agreement no. 633962). G.L.P. acknowledges the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and the PhD scholarship 434655 from CONACyT.

Author contributions

J.U. and G.L.P. designed ReProMin. G.L.P. developed computational methods and performed data analysis. G.L.P. and J.S.M.H. carried out experiments. G.L.P., J.K. and J.I.J. analyzed flow cytometry experiments, isocost lines and violacein production. J.U. supervised the study. J.U., G.L.P. and J.I.J. wrote the manuscript.

Competing interests

J.U. and G.L.P. are inventors in a MX patent application filed by UNAM.

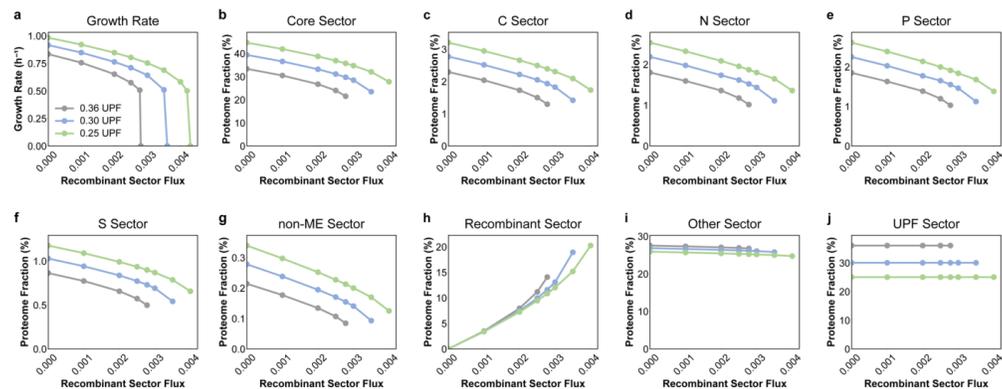
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41589-020-0593-y>.

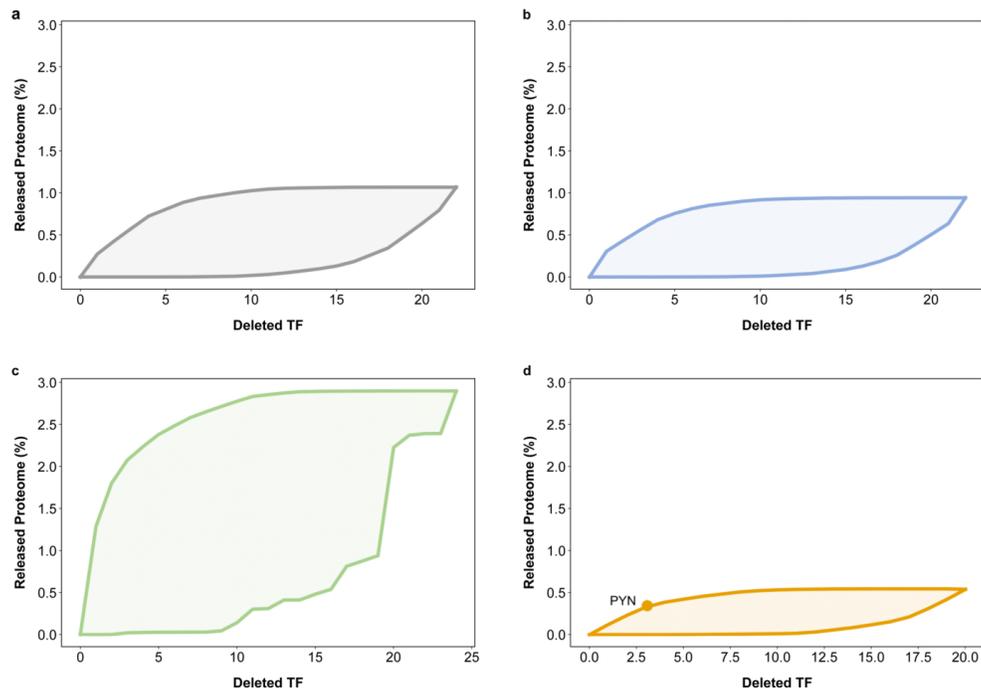
Supplementary information is available for this paper at <https://doi.org/10.1038/s41589-020-0593-y>.

Correspondence and requests for materials should be addressed to J.U.

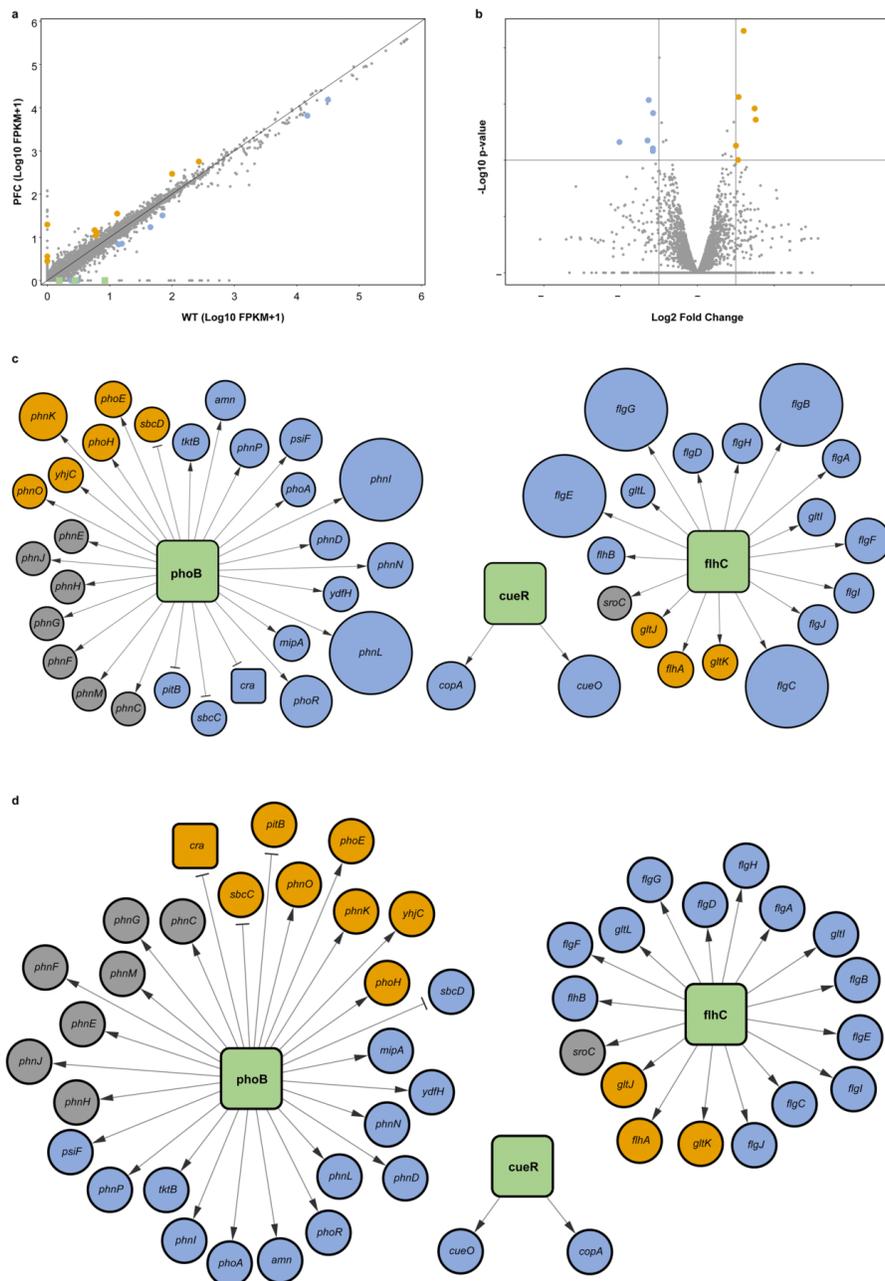
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | ME-model simulations and proteome sector response to reducing the unmodeled protein fraction (UPF). The ME-model *iJL1678b*-ME was used to simulate the effect of the reduction of the UPF and different expression levels of an unused recombinant model protein (GFP) (see methods). Similar to the maintenance energy coefficient, the hedging proteome and other non-growth related (thus not modeled) functions are accounted for in ME-models as a part of the UPF. Each panel shows **a**, growth rate and the corresponding fraction of each proteome sector **b**, core sector and the alternative element dependent sector: **c**, the carbon sector **d**, the nitrogen sector **e**, the phosphate sector **f**, the sulphur sector **g**, the non-ME sector **h**, the recombinant sector, comprised by the maximum attained GFP expression, **i**, the other sector (non-classified) and **j**, the UPF sector. The simulation shows an increased availability of cellular resources for recombinant protein production by reducing the UPF.

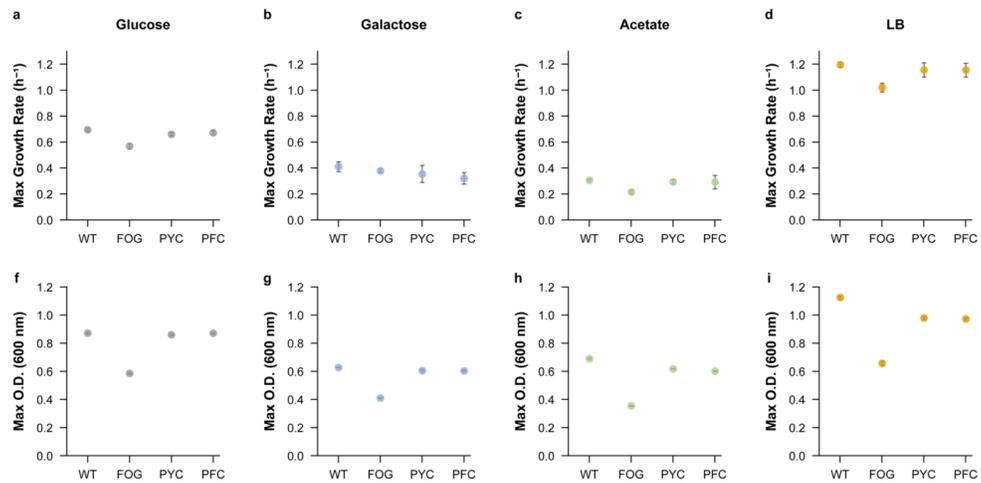


Extended Data Fig. 2 | ReProMin proteome liberation landscapes corresponding to the UT case. Potential proteome liberation landscape corresponding to **a**, Galactose, **b**, Acetate, **c**, Glycerol + casAA and **d**, Rich Medium (LB).

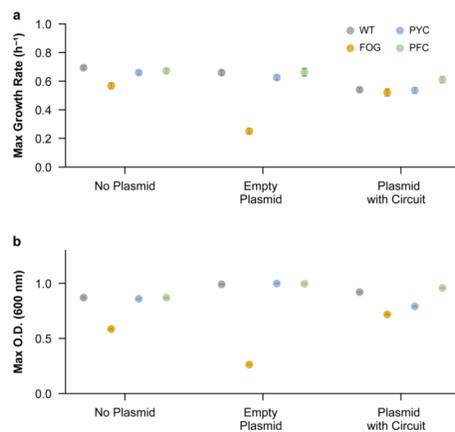


Extended Data Fig. 4 | See next page for caption.

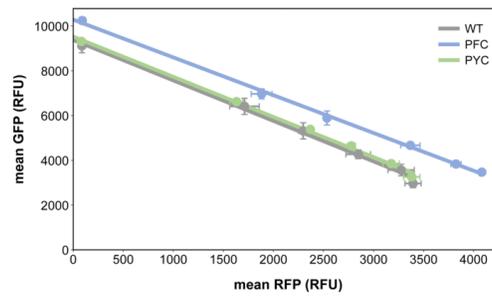
Extended Data Fig. 4 | Transcriptomic analysis of the UT case designed strain. **a.** Correlation plot for PFC and WT strains transcripts. Green squares represent the three deleted TFs. **b.** Volcano plot showing differential gene expression. In both cases, statistically significant genes are highlighted (blue – downregulated, yellow – upregulated) (\log_2 Fold Change ≥ 1 or ≤ -1 , $P \leq 0.01$, $n = 2$). **c.** Integration of transcriptomics with computational tool predictions. The size of the circle corresponds to the fold change of each target (the largest circles represent fully silenced genes), in all cases blue circles represent targets releasing resources (down regulated), yellow circles represent targets generating burden (upregulated) and grey circles targets that were not found expressed. **d.** Accuracy of computational tool predictions based on RNAseq data. Yellow circles represent wrong predictions, blue circles represent accurate predictions and grey circles represent unmapped predictions (expression was not detected).



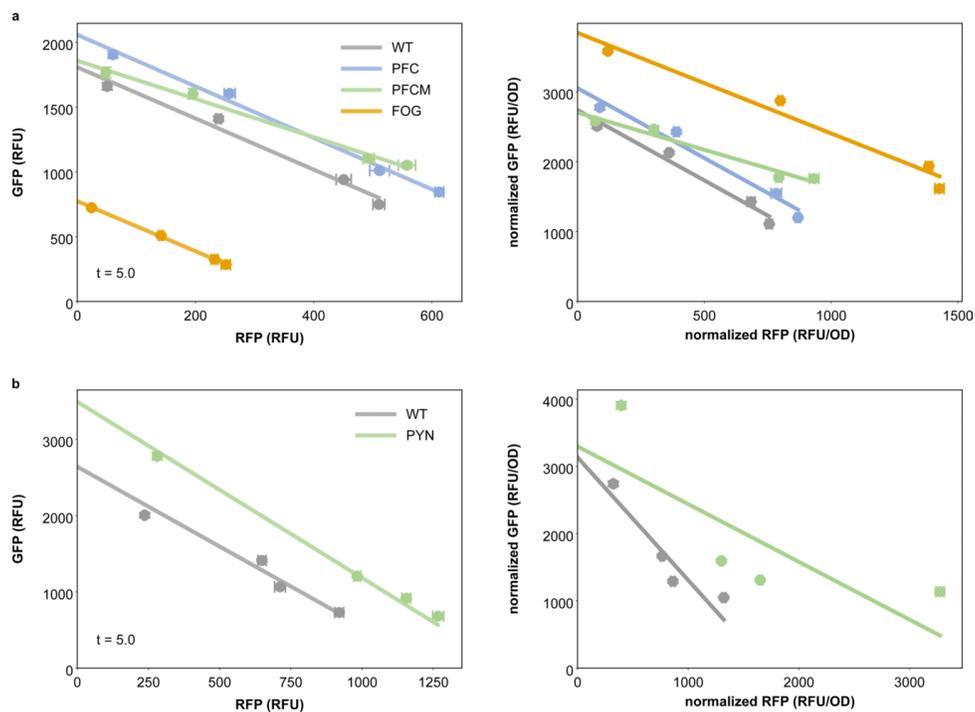
Extended Data Fig. 5 | Phenotypic evaluation generated strains based on glucose ReProMin predictions (UT and ST cases) and control. Growth on different carbon source supplemented M9 medium and rich medium (LB). **a-d**, shows max growth rate and **e-h**, shows max O.D. Points represent the Gaussian fitted value \pm 2 s.d. for $n=9$.



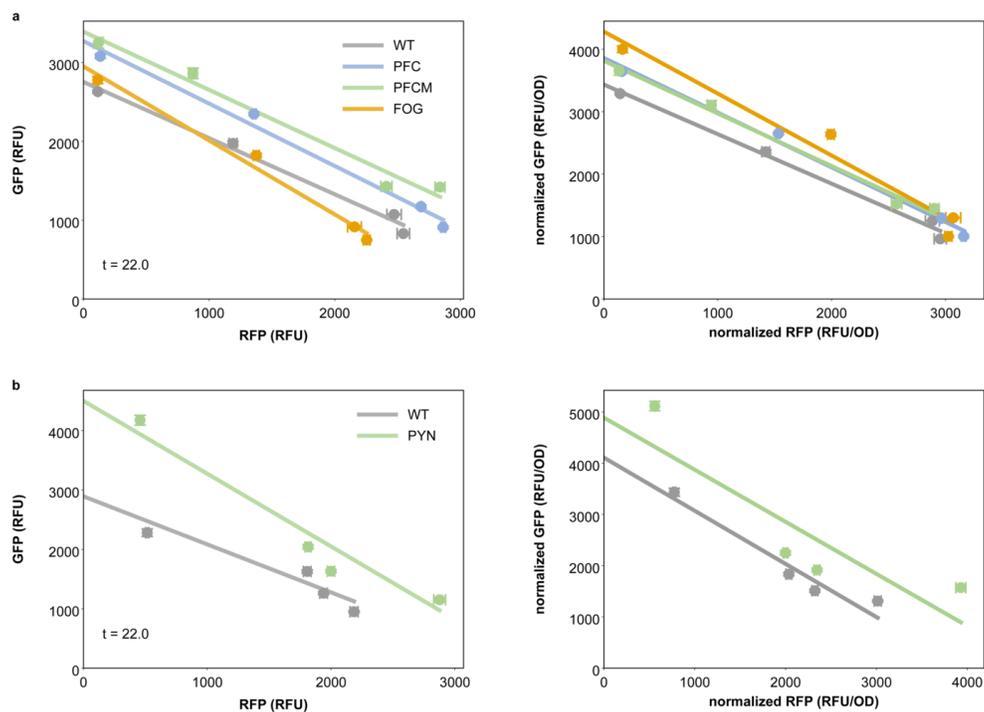
Extended Data Fig. 6 | Metabolic burden evaluation of strains based on glucose ReProMin predictions (UT and ST cases) and control. Metabolic burden while carrying empty, circuit plasmid and induced circuit plasmid, **a**, shows max growth rate and **b**, shows max O.D. Points represent the Gaussian fitted value \pm 2 s.d. for $n=9$.



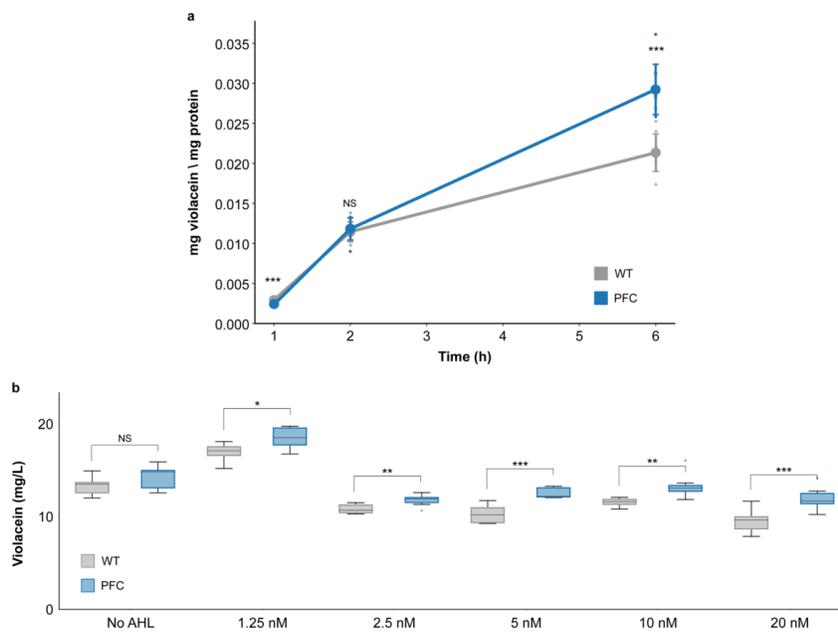
Extended Data Fig. 7 | Synthetic circuit characterization in glucose M9 medium. Isocost lines showing mean fluorescence per cell measured by flow cytometry during balanced growth (~5 h). Points represent the Gaussian fitted fluorescence value \pm 2 s.d. for $n=9$ of red reporter (x axis) plotted against the green reporter (y axis) in an increasing inducer concentration (0, 2.5, 5, 20 nM AHL). A linear regression was used to fit the points to a line.



Extended Data Fig. 8 | Isocost lines during balanced growth (-5h). Isocost lines of the generated mutant strains for two growth conditions: **a**, Glucose M9 medium and **b**, Rich medium. Left: absolute fluorescence, Right: normalized fluorescence. Points represent the Gaussian fitted fluorescence value \pm 2 s.d. for $n=9$ of red reporter (x axis) plotted against the green reporter (y axis) in an increasing inducer concentration (0, 2.5, 5, 20 nM AHL). A linear regression was used to fit the points to a line.



Extended Data Fig. 9 | Isocost lines during stationary phase (-22 hrs). Isocost lines of the generated mutant strains for two growth conditions: **a**, Glucose M9 medium and **b**, Rich medium. Left: absolute fluorescence, Right: normalized fluorescence. Points represent the Gaussian fitted fluorescence value ± 2 s.d. for $n=9$ of red reporter (x axis) plotted against the green reporter (y axis) in an increasing inducer concentration (0, 2.5, 5, 20 nM AHL). A linear regression was used to fit the points to a line.



Extended Data Fig. 10 | Violacein production characterization. **a**, Protein normalized violacein production using 2 g/L tryptophan in the presence of AHL (20 nM) (mean \pm s.d., $n=9$). **b**, Total violacein production without adding tryptophan after 24 h in the presence of increasing inducer (AHL) concentrations (mean \pm s.d., $n=9$). Asterisks *, ** and *** denote significant differences between WT and PFC using a two-tailed unpaired Student's *t*-test. The following *P* values were obtained for normalized violacein production: 1 h, $P=0.0003$; 2 h, $P=0.5647$; 6 h, $P<0.0001$. The following *P* values were obtained for violacein production with different AHL concentrations: No AHL, $P=0.0599$; 1.25 nM, $P=0.0146$; 2.5 nM, $P=0.0021$; 5 nM, $P<0.0001$; 10 nM, $P=0.0014$; 20 nM, $P=0.0005$.

Bibliografía

1. Way, J. C., Collins, J. J., Keasling, J. D. & Silver, P. A. Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go. *Cell* **157**, 151–161 (2014).
2. Ishihama, A. Prokaryotic genome regulation: Multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiology Reviews* vol. 34 628–645 (2010).
3. Ulrich, L. E. & Zhulin, I. B. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res.* **38**, D401-7 (2010).
4. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5**, 826–837 (2004).
5. Isalan, M. *et al.* Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840–845 (2008).
6. Mitchell, A. *et al.* Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220–224 (2009).
7. Tagkopoulos, I., Liu, Y.-C. & Tavazoie, S. Predictive behavior within microbial genetic networks. *Science* **320**, 1313–1317 (2008).
8. Acar, M., Mettetal, J. T. & Van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**, 471–475 (2008).
9. Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C. & Rainey, P. B. Experimental evolution of bet hedging. *Nature* **462**, 90–93 (2009).
10. Utrilla, J. *et al.* Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution. *Cell Syst.* **2**, 260–71 (2016).
11. Orth, J. D. & Palsson, B. Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions. *BMC Syst Biol* **6**, 30 (2012).
12. O'Brien, E. J., Lerman, J. a, Chang, R. L., Hyduke, D. R. & Palsson, B. Ø. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9**, 693 (2013).
13. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science (80-.)*. **351**, aad6253–aad6253 (2016).
14. Balikó, G. *et al.* Rational efforts to streamline the escherichia coli genome. in *Synthetic Biology: Parts, Devices and Applications* 49–80 (Wiley-VCH Verlag GmbH & Co. KGaA, 2018).

15. Scott, M., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of Cell Growth Origins and Consequences. *Science* (80-.). **330**, 1099–1102 (2010).
16. Hui, S. *et al.* Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784 (2015).
17. Nikolados, E.-M., Weiße, A. Y., Ceroni, F. & Oyarzún, D. A. Growth Defects and Loss-of-Function in Synthetic Gene Circuits. *ACS Synth. Biol.* acssynbio.8b00531 (2019) doi:10.1021/acssynbio.8b00531.
18. Glasscock, C. J., Lucks, J. B. & DeLisa, M. P. Engineered Protein Machines: Emergent Tools for Synthetic Biology. *Cell Chem. Biol.* **23**, 45–56 (2016).
19. Monk, J. M. *et al.* iML1515, a knowledgebase that computes Escherichia coli traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
20. O'Brien, E. J. *et al.* Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments. *PLOS Comput. Biol.* **12**, e1004998 (2016).
21. Yang, L. *et al.* Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Sci. Rep.* **6**, 36734 (2016).
22. McCloskey, D., Palsson, B. Ø. & Feist, A. M. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol. Syst. Biol.* **9**, 661 (2013).
23. Lloyd, C. J. *et al.* COBRAME: A computational framework for genome-scale models of metabolism and gene expression. *PLOS Comput. Biol.* **14**, e1006302 (2018).
24. Santos-Zavaleta, A. *et al.* RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).
25. Schmidt, A. *et al.* The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.* **34**, 104–110 (2015).
26. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**, 64–68 (2002).
27. Kim, J., Darlington, A., Salvador, M., Utrilla, J. & Jiménez, J. I. Trade-offs between gene expression, growth and phenotypic diversity in microbial populations. *Curr. Opin. Biotechnol.* **62**, 29–37 (2020).
28. Turner, P. C. *et al.* YqhC regulates transcription of the adjacent Escherichia coli genes yqhD and dkgA that are involved in furfural tolerance. *J. Ind. Microbiol. Biotechnol.* **38**, 431–439 (2011).
29. Yakovleva, G. M., Kim, S. K. & Wanner, B. L. Phosphate-independent expression of the carbon-phosphorus lyase activity of Escherichia coli. *Appl.*

- Microbiol. Biotechnol.* **49**, 573–8 (1998).
30. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophys. J.* **109**, 639–46 (2015).
 31. Bienick, M. S. *et al.* The Interrelationship between Promoter Strength, Gene Expression, and Growth Rate. *PLoS One* **9**, e109105 (2014).
 32. Durán, N. *et al.* Advances in *Chromobacterium violaceum* and properties of violacein-Its main secondary metabolite: A review. *Biotechnol. Adv.* **34**, 1030–1045 (2016).
 33. Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, 695 (2018).
 34. Baumstark, R. *et al.* The propagation of perturbations in rewired bacterial gene networks. *Nat. Commun.* **6**, 10105 (2015).
 35. Fang, X. *et al.* Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proc. Natl. Acad. Sci.* 201702581 (2017) doi:10.1073/pnas.1702581114.
 36. Ibarra-Arellano, M. A., Campos-González, A. I., Treviño-Quintanilla, L. G., Tauch, A. & Freyre-González, J. A. Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria. *Database* **2016**, baw089 (2016).
 37. Sastry, A. V. *et al.* The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **10**, 1–14 (2019).
 38. Hidalgo, D. & Utrilla, J. Resource Allocation Principles and Minimal Cell Design. in *Minimal Cells: Design, Construction, Biotechnological Applications* 211–230 (Springer International Publishing, 2020). doi:10.1007/978-3-030-31897-0_8.
 39. Couto, J. M., McGarrity, A., Russell, J. & Sloan, W. T. The effect of metabolic stress on genome stability of a synthetic biology chassis *Escherichia coli* K12 strain. *Microb. Cell Fact.* **17**, 8 (2018).
 40. Choe, D. *et al.* Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* **10**, 935 (2019).
 41. Posfai, G. *et al.* Emergent Properties of Reduced-Genome *Escherichia coli*. *Science (80-.)*. **312**, 1044–1046 (2006).
 42. Lastiri-Pancardo, G., Mercado-Hernández, J. S., Kim, J., Jiménez, J. I. & Utrilla, J. A quantitative method for proteome reallocation using minimal regulatory interventions. *Nat. Chem. Biol.* **16**, 1026–1033 (2020).
 43. Price, M. N., Wetmore, K. M., Deutschbauer, A. M. & Arkin, A. P. A Comparison of the Costs and Benefits of Bacterial Gene Expression. *PLoS*

- One* **11**, e0164314 (2016).
44. Yang, L. *et al.* Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data. *Proc. Natl. Acad. Sci.* **112**, 10810–10815 (2015).
 45. Kato, J.-I. & Hashimoto, M. Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* **3**, 132 (2007).
 46. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
 47. Yamamoto, N. *et al.* Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* **5**, 335 (2009).
 48. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
 49. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
 50. Miller, J. H. *A short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria.* (Cold Spring Harbor Laboratory Press, 1992).
 51. St-Pierre, F. *et al.* One-Step Cloning and Chromosomal Integration of DNA. *ACS Synth. Biol.* **2**, 537–541 (2013).
 52. Swain, P. S. *et al.* Inferring time derivatives including cell growth rates using Gaussian processes. *Nat. Commun.* **7**, 13766 (2016).